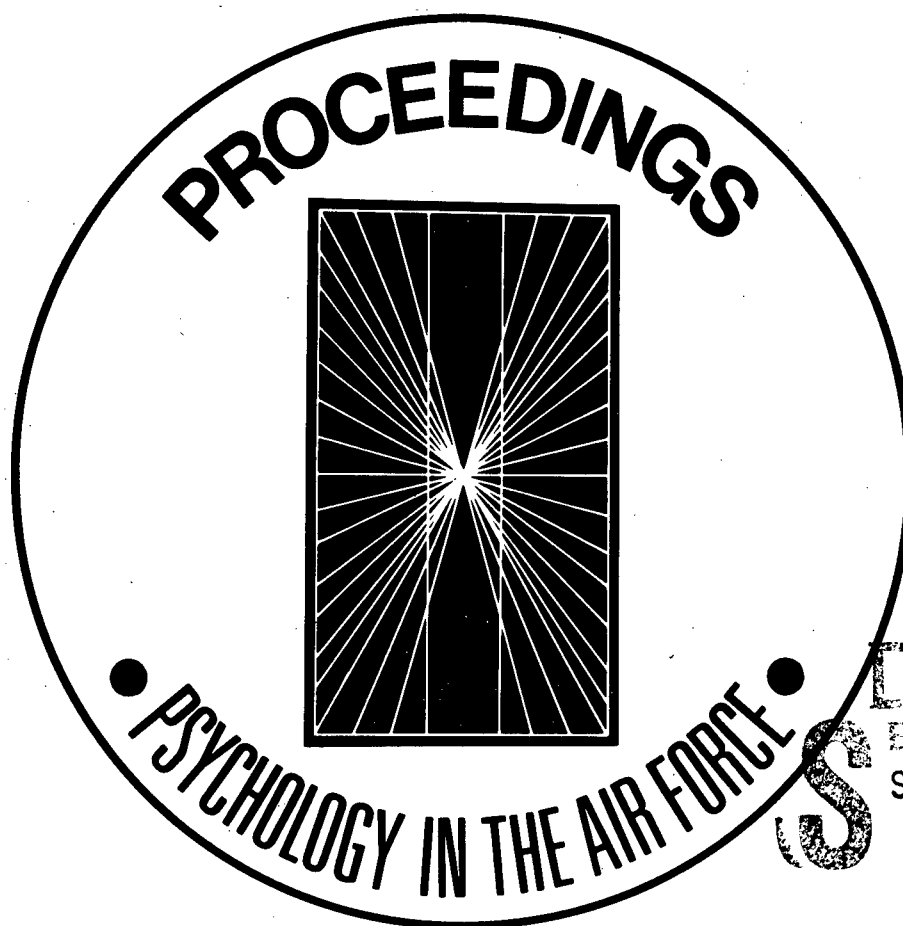


1



DTIC
ELECTE
SEP 24 1991
S D D

19970828 083

3rd ANNUAL SYMPOSIUM

APRIL 1972

This document has been approved
for public release and sale; its
distribution is unlimited.

SPONSOR: USAF ACADEMY DEPARTMENT OF LIFE AND BEHAVIORAL SCIENCES


919 23 071

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Hq USAFA (Faculty Secretariat) USAF Academy, Colorado 80840		2a. REPORT SECURITY CLASSIFICATION Unclassified	
		2b. GROUP n/a	
3. REPORT TITLE Proceedings, Psychology in the Air Force, 3rd Annual Symposium			
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Symposium proceedings, held three inclusive days 18-20 April 1972			
5. AUTHOR(S) (First name, middle initial, last name) Hal W. Hendrick, Major, USAF Robert B. Tebbs, Major, USAF Proceedings Editors Lloyd R. Chason, Major, USAF			
6. REPORT DATE April 1972		7a. TOTAL NO. OF PAGES 315	7b. NO. OF REFS n/a
8a. CONTRACT OR GRANT NO. USAF Academy Authorized Publication		9a. ORIGINATOR'S REPORT NUMBER(S) USAFA Research Report # 72-10	
b. PROJECT NO. n/a			
c. n/a		9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
d. n/a		n/a	
10. DISTRIBUTION STATEMENT These documents are approved for public release, distribution unlimited, in accordance with paragraph, IVA, AFR 80-45			
11. SUPPLEMENTARY NOTES n/a		12. SPONSORING MILITARY ACTIVITY Dept of Life & Behavioral Sciences USAF Academy, Colorado 80840	
13. ABSTRACT These proceedings encompass papers presented dealing with a wide range of psychology and research in all military establishments in the DOD: Personnel, Training, Retraining, Human Factors, Human Engineering, Drugs, Alcohol, Radiation, Attitudes, Race Relations, Motivation, Systems Engineering, Anthropometric-Aircraft Research, Performance, Perception, Behavior, Operations Research and All Volunteer Force			
 DTIC QUALITY INSPECTED UNANNOUNCED		Accession For	
		NTIS CRA&I <input checked="" type="checkbox"/>	
		DTIC TAB <input type="checkbox"/>	
		Unannounced <input type="checkbox"/>	
		Justification	
		By	
		Distribution /	
		Availability Codes	
		Dist	Avail and/or Special
		A-1	

DD FORM 1473
1 NOV 65

Unclassified

Security Classification

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Psychology						
Symposium						
Air Force						
Military						
Personnel						
Training						
Retraining						
Human Factors						
Human Engineering						
Drugs						
Alcohol						
Radiation						
Attitudes						
Race Relations						
Motivation						
Systems Engineering						
Anthropometric-Aircraft Research						
Performance						
Perception						
Behavior						
Operations Research						
All Volunteer Force						

Unclassified

Security Classification

THIRD ANNUAL SYMPOSIUM
PSYCHOLOGY IN THE AIR FORCE

18 April - 20 April 1972

*Department of Life and Behavioral Sciences
United States Air Force Academy
Colorado Springs, Colorado*

*Professor and Head, Department of Life and Behavioral Sciences
USAF Academy*

Colonel Peter B. Carter

Symposium Chairman

Major Robert B. Tebbs

Assistant Symposium Chairman

Major Lawrence F. Sharp

edited by:

Hal W. Hendrick, Major, USAF
Associate Professor and Director of Behavioral Research
Robert B. Tebbs, Major, USAF
Associate Professor and Director of the Experimental Division
Lloyd R. Chason, Major, USAF
Associate Professor and Director of Research and Laboratories

Papers presented by:

Norman M. Abrahams	Patricia A. Knoop
Douglas Adair	Jefferson M. Koonce
William E. Alley	James C. Marsh
James D. Basinger	William P. Mockovak
Robert W. Bauer	Cecil J. Mullins
John M. Bermudez	Idell Neumann
William E. Beusse	Victor N. Ogilvie
Gene A. Berry	Michael J. Ozerkevich
Robert F. Boldt	Gordon A. Parrish
Gerald P. Chubb	Dirk C. Prather
Bertram W. Cream	Daniel L. Rickett
Gary J. Echternacht	Lawrence A. Scanlan
Richard S. Elster	Debbie J. Seifert
William L. Foley	Wayne S. Sellman
Nancy Frost	Lawrence F. Sharp
William H. Githens	Gregory D. Smith
Harvey G. Gregoire	James F. Smith
Don R. Gum	Mark L. Stein
Israel M. Guterman	Arthur B. Sweney
Hal W. Hendrick	Robert B. Tebbs
William H. Hendrix	Edmund D. Thomas
Russell J. Hibler	Patricia J. Thomas
Charles R. Holloman	Horace H. Valverde
Earl Hunt	Bart M. Vitola
Larry H. Ingraham	Donnell L. Washington
Robert S. Jacobs	John W. Williams
T. O. Jacobs	James A. Wise
James P. Jenkins	Joseph D. Young
Richard S. Jensen	Anchard F. Zeller
Anna May Kelleher	

NOTE:

The opinions expressed in this volume are those of the individual authors and do not necessarily reflect official policy either of the Army, Navy or Air Force or of any government organization in which the author may be serving.

Trade names of materials or products of commercial or non-government organizations are cited only where essential to precision in describing research procedures or evaluation of results. Their use does not constitute official indorsement or approval of the use of such commercial hardware or software.

These proceedings are approved for public release, distribution unlimited, in accordance with paragraph IVA, AFR 80-45, 26 March 1971.

Third Annual Symposium
Psychology in the Air Force

TABLE OF CONTENTS

<i>Keynote Address</i> -- Mr. Donald L. Miller, Deputy Assistant Secretary of Defense, Equal Opportunity - Manpower & Reserve Affairs	1
SESSION I	
Social Psychology	
Chairman: Major Robert B. Tebbs	
<i>Descriptive Modeling of Non-frequentistic Subjective Probability</i> Lt. William P. Mockovak	13
<i>Television Distortion and Attitude Change -- A Backlash Effect</i> Captain Michael J. Ozerkevich	19
<i>Divorce and Family Dissolution in a Military Environment</i> Lt Col John W. Williams	23
<i>Strategies for Conducting Mission-Oriented Research in Military Organizations</i> Dr. Arthur B. Sweney	37
<i>Factors Affecting Group Decision-Making</i> Major Hal W. Hendrick	43
Training - Minority Groups	
Chairman: Major Hal W. Hendrick	
<i>Imagination as a Flight Simulator</i> Major Dirk C. Prather	59
<i>Navy Fighter Pilots' Responses to Paper-and-Pencil Simulations for Air-to-Air Combat Situations</i> Dr. Richard S. Elster	63

*Transfer and Stress: A Comparison of Student Performances
Resulting from Trial-and-Error and Prompting and Feedback
Instructional Techniques in the Learning of a Perceptual
Skill* 71
Captain Gene A. Berry

*An Investigation of Possible Test Bias in the Navy Basic
Test Battery* 77
Patricia J. Thomas and Edmund D. Thomas

SESSION II

Training

Chairman: Major Dirk C. Prather

*Service Test/Evaluation of Multimedia CDC 60300, Vehicle
Operator/Dispatcher* 87
Lt Russell J. Hibler

*Development of an Advanced Training Research Simulation
System* 95
Don R. Gum

*Application of Advanced Simulation in Undergraduate Pilot
Training (ASUPT) Research Facility to Pilot Training
Problems* 105
Mr. James F. Smith

*Development and Evaluation of Two Functional Part-Task
Trainers* 113
Mr. Horace Valverde and Mr. Bertram W. Cream

Human Factors (Dual Session)

Chairman: Captain Frederick V. Malmstrom

*Effects of Electronically Produced Airborne Noise on
Psychophysical Performance of Military Tasks* 125
Mr. James P. Jenkins

*Can You Reach the Controls? A Cockpit Anthropometric
Survey* 135
LCDR Harvey G. Gregoire

*A TSD Determination of the Two-Point Supraliminal DLS on the
Dorsal Forearm, the Anterior Thigh, and the Back of the Hand* 145
Captain Donnell M. Washington

<i>Performance as a Function of Task Difficulty in a Crespi Reversal Situation</i> Major Lawrence F. Sharp	155
<i>Private Motor Vehicle Accident Rates of USAF Pilots vs Nonpilot Officers</i> Dr. Anchar F. Zeller	163
Training (Dual Session)	
Chairman: Major Joseph K. Jarboe	
<i>Confidence Tests as Diagnostic Aids in Technical Training</i> Captain Wayne S. Sellman	171
<i>Preliminary Studies of a Betting Model in Achievement Testing to Predict Troubleshooting Success in the Lab</i> Ms Anna May Kelleher	179
<i>Reduction of Automated Readability Index Calculation Time</i> Captain Joseph D. Young	185
<i>Utilization, Success and Benefits of the USAFI GED Program</i> Lt William E. Beusse	191
Drugs and Rehabilitation	
Chairman: Captain Donnell L. Washington	
<i>The Vietnam Heroin Epidemic: A Descriptive Profile of the Army Returnee</i> Captain Larry H. Ingraham	199
<i>Comparison of Personal Characteristics of Identified Drug Users with Non-Users</i> Mr. Bart M. Vitola	205
<i>Attitude Change Among Selected Air Force Prisoners</i> Captain Victor N. Ogilvie	213
<i>Military Humanism: Some Further Adventures in Rehabilitation</i> Colonel Ogden Brown, Jr.	223

SESSION III

Human Factors (OAR Sponsored Research)

Chairman: Major Lawrence F. Sharp

<i>Predictive Validity of Ground-based Flight Checks</i>	233
Major Jefferson M. Koonce	
<i>Motion Cues as a Factor in Simulator and Airborne Evaluation of Flight Director Displays</i>	239
Mr. Robert S. Jacobs	
<i>Spatial and Temporal Aspects of Perception with Visually Time-Compressed Displays</i>	249
Mr. Lawrence A. Scanlan	
<i>An Assessment of Symbolic Area Navigation Display Variables</i>	257
Mr. Richard S. Jensen	
<i>Evidence for a Process Model of Intelligence</i>	265
Dr. Earl B. Hunt	

Human Factors

Chairman: Captain Richard L. Hughes

<i>Perceived Desirability of Assignment as a Personnel Subsystem Manager</i>	275
Captain William H. Hendrix	
<i>Army Aircraft Survivability Human Factors</i>	281
Dr. Robert W. Bauer	
<i>A Psychologist's Input to Operations Research</i>	287
Dr. Gerald P. Chubb	
<i>Determinants of the Post-Arousal Performance Decrement: Implications for Research and Applied Psychology</i>	293
Major Robert B. Tebbs	
<i>Personality Correlates of Left-Handers</i>	299
Lt Mark L. Stein	

All-Volunteer Force

Chairman: Captain John M. Bermudez

<i>Prediction of Disenrollment and Military Aptitude at the Naval Academy with the Strong Vocational Interest Blank</i> Dr. Norman M. Abrahams	309
<i>A Preliminary Analysis of the 1970 OTS Applicant Pool by Draft Vulnerability Category</i> Mr. William E. Alley	317
<i>An Analysis of the Impact of VOLAR (Volunteer Army) Actions at Fort Benning</i> Dr. T. O. Jacobs	327

PANEL DISCUSSIONS*

SESSION I

Minority Group Relations

Chairman: *Major Lloyd R. Chason, USAF Academy (DFLS)*

Participants:

*Harry H. Jones, Dept. of the Air Force (AFHRL)
LCDR Gordon E. Fisher, Dept. of the Navy
Colonel H. W. Brooks, Jr., Dept. of the Army
Major J. F. Borus, Dept. of the Army*

Discussant: *Lt Col H. E. Mathews, Dept. of the Air Force*

SESSION II

Drug and Alcohol Use-Abuse

Chairman: *Major Eugene H. Galluscio, USAF Academy (DFLS)*

Participants:

*Major Brian Doyle, Dept. of the Army
CDR John Brennan, Dept. of the Navy
Lt Col James T. Ronaghan, Dept. of the Air Force
(AFA)*

Discussant: *Lt Col W. L. Williams, Dept. of the Air Force
(DOD)*

SESSION III

Impact of the All Volunteer Force Concept

Chairman: *Major Jock C. H. Schwank, USAF Academy (DFLS)*

Participants:

*Captain Joseph Siefert, Dept. of the Army
Lt Col Hal Van Meter, Dept. of the Army
CDR A. G. Lacey, Dept. of the Navy*

Discussant: *Lt Col G. T. Carpenter, Royal Military College
of Canada*

*Panel Discussions on the topics listed were held as part of the Symposium, however, the Proceedings do not contain summaries of the content of these discussions. For those who desire information about the panels, it is suggested contact be made with the participants concerned.

SESSION I

Social Psychology

Training - Minority Groups

Social Psychology

CHAIRMAN: Major Robert B. Tebbs

DESCRIPTIVE MODELING OF NON-FREQUENTISTIC

SUBJECTIVE PROBABILITY

William P. Mockovak and James A. Wise

Ohio State University - AFIT

In a task subjects posed as medical researchers who must inspect different virus (schematic drawings) and then subjectively estimate their probability of being cancerous. Having done this the subjects then had to decide what further tests would be performed on the virus. A geometric model was presented in which the different virus were represented as points in an n-dimensional space. Subjective probabilities were therefore viewed as a form of similarities data but differed insofar as the distance function needed to describe them. The predictive accuracy of the model was also tested.

Descriptive models of risky decision-making are partly based on an individual's "degree of belief" in the occurrence of certain outcomes. This notion has been formalized by de Finetti (1937) and others as subjective probability. Often, judgments of events involving subjective probabilities are made under conditions where relative frequencies are completely undefined or inappropriate. In these situations then, how does a person arrive at a subjective probability estimate which represents his "degree of belief"? Wise (1970) devised and empirically examined a model of non-frequentistic subjective probability.

Generally, the model states that subjective probability estimates can be based on the relative similarity of stimuli or situations. In the model, relative similarities are represented by a configuration of interstimulus distances in a cognitive space. A measure of the subjective probability that any stimulus(i) belongs in the same class as a known stimulus(i) is then defined as:

$$P(i) = \frac{1/d_{it}}{\sum_{j=1}^r 1/d_{jt}}$$

where there are r alternatives, and d is the appropriate cognitive distance function.

This hypothesis has been supported in previous experiments (Wise, 1970) which demonstrated that subjective probability estimates possess strong consistency properties predicted by the model. In that study however, there was no independent assessment of the appropriate cognitive distance function.

The present paper meets this need by utilizing a multidimensional scaling analysis (Kruskal, 1964) to independently assess the appropriate cognitive distance function for a set of similarity judgments. Then, subjective probability estimates are collected for related sets of stimuli and the cognitive distance function from the MDSA is used as input to the predictive model.

Method

The experiment was conducted in two parts. In the first part Ss made dissimilarity ratings of pairs of schematic pictures of viruses by drawing a straight line on blank sheets of paper in a response booklet. The length of the line was indicative of the relative dissimilarity of the two viruses being judged. That is, the more dissimilar two viruses were, the longer the line should be that was drawn to represent their difference.

In the second part of the experiment, the Ss saw other schematic pictures of viruses displayed in triads. In every case the top virus was labeled as cancerous, and the S had to estimate his subjective probability that each of the other viruses was cancerous, if one of them also had to belong in that class. There was one set of nine base viruses from which two other sets were derived by the application of certain selected transformations on the stimulus dimensions. These transformations altered stimulus appearance, but left subjective probability invariant with respect to the model being tested. The use of such transformations was suggested by Wise (1971), and represent an effort to empirically test de Finetti's (1937) "exchangeability" requirement for subjective probabilities.

Subjects

Sixty-three male Ss from an introductory psychology course at Ohio State University participated in groups of 1-3.

Stimuli

The stimuli were 27 schematic drawings of viruses which were constructed on the basis of three orthogonal dimensions. The first dimension (X) was the length of a virus' body fibrils, the second dimension (Y) was the length of a virus' body side, and the third dimension (Z) was the distance between two "nucleoles" in a virus' body. There were three basic groups of viruses (nine viruses in a group), a base group, a rotation group, and a dilatation group. The base group was constructed to give a wide range of subjective probability estimates, and the rotation and dilatation groups were obtained from the base group through the application of selected transformations on the original stimulus dimensions (X, Y, Z). The dilatation transformation was the same as that used by Wise (1970), but the rotation transformation involved rotating the XY axis in the

Cartesian coordinate system two degrees and the XZ axis ten degrees. Further organization of the stimuli parallels that of Wise (1970).

Procedure

The Ss were asked to imagine themselves as medical researchers in a laboratory involved in studying different kinds of viruses. It was explained to them that in order to save time and money, the laboratory wanted to know if researchers could run the proper test on a virus using only visual information obtained from schematic pictures of the viruses. In both parts of the experiment, the Ss were specifically instructed to base their decisions equally on each of the three dimensions X, Y, and Z; also, they were told not to use strategies or mathematical formulae they might have learned in mathematics courses. It was impressed upon them that there were no "right" answers to what they were doing, and therefore they should rely on their "intuition" as much as possible.

After the dissimilarity judgments were made, it was further explained to the Ss that in the second part of the experiment, the more two viruses were similar in appearance, the more likely they were of being the same type. Therefore, for each of the triads, the bottom virus that most resembled the top virus had the greater chance of being cancerous. Ss recorded their subjective probability estimates by marking a scale, graduated in .01 units, between 0.0 and 1.0.

Results

An analysis of the MDSA indicated that the 3-dimensional, Euclidean solution provided the best fit for the dissimilarity data. Since the Kruskal program provides coordinates for the final configuration of points in a solution space, it was possible to rank order the coordinate values for a given dimension using the Euclidean solution, and compare it with the rank order of the original coordinate values to determine whether or not the Ss were using the original specified dimensions (X, Y, Z) for their dissimilarity judgments. Use of Kendall's rank correlation coefficient revealed that all three groups of Ss used the Y dimension, ($p < .05$) but only groups 1 and 2 used the X dimension as well, ($p < .05$) The Z dimension was never reproduced by the MDSA.

From the original dimensions (X, Y, Z) it was possible to calculate the Euclidean distances between viruses. These distances were then substituted into Wise's (1970) probability measure to yield predicted subjective probabilities for the triadic judgments in part two of the experiment. The Ss' actual median subjective probability estimates were then compared with these predicted values. It was found that the measure had high predictive accuracy resulting in a Pearson product moment correlation of $r = .83$ and $b = 1.07$.

The consistency of the Euclidean model is describable in terms of transformations of the model that leave certain probability predictions invariant. There were three such transformations in this experiment: reflection, rotation, and dilatation. Reflection compared what should be equivalent subjective probability estimates within all the triads (see Wise, 1970). Consistency was good, $r = .87$ and $\hat{B} = .99$ (corrected in accordance with Isaac, 1970). The consistency requirements of the Euclidean model also require that for the three triads of viruses in the base group, the subjective probability estimates should not change for the corresponding triads in the rotation and dilatation groups. The comparison between the base and rotated group resulted in $r = .86$ and $\hat{B} = .72$; whereas, the comparison between the base and dilated group resulted in $r = .96$ and $\hat{B} = .99$.

Variability of the S_s' estimates was calculated by the semi-interquartile range Q , and was generally less than .10.

Discussion

The present study served as a further test of the model proposed by Wise (1970b). In this case, however, an independent assessment of the appropriate cognitive distance function was made possible through the use of a multidimensional scaling analysis. It is interesting to note that a Euclidean spatial model was found to be more acceptable than a City-Block. This was true even though the stimulus dimensions were perceptually distinct, since other studies have argued that such stimuli result in a City-Block solution for similarities data (Hyman & Well, 1967, 1968). Whether the City-Block or Euclidean model is more basic for similarities data is a controversy which has not been successfully resolved (Hoijer, 1969), and it was of interest in this study only because the different distance measures predict different subjective probability estimates when they are used in Wise's model.

Besides the use of different stimuli, this study also introduced another transformation in addition to the dilatation transformation present in Wise's experiment. The rotation transformation introduced had a two-fold purpose. First of all, along with the dilatation transformation, it served as an empirical test of de Finetti's (1937) "exchangeability" requirement for probability. The exchangeability requirement as it relates to personal probabilities has been explained in detail elsewhere (Wise, 1971), but generally it refers to the allowable transformations that can be imposed on a cognitive space while leaving subjective probability invariant. The other purpose of the rotation transformation was to further differentiate between the City-Block and Euclidean distance functions. The values of rotation were chosen to produce the largest possible discrepancy between the predicted subjective probability estimates of the two measures.

The results of this study for the dilatation transformation were found to be consistent with those of Wise (1970). In a Euclidean spatial model, the rotation transformation is also an allowable transformation since the Euclidean model possesses rotational invariance. The lower consistency that resulted under the rotation transformation could have been the result of several factors. First of all, the possibility exists that a continuum of combining rules is applicable (Hyman & Well, 1968). The influence of these other spatial models would have the most effect in the rotation transformation since they do not possess rotational invariance. However, a more likely reason is uncontrolled proportionality changes between the stimulus dimensions under the rotation transformation which resulted in a built-in bias in the S_s' weighting of the different stimulus dimensions. This bias caused the Y dimension to be weighted more than the Z, the X to be weighted more than the Z, and the X more than the Y. Feedback from the S_s at the end of the experiment indicated that in their opinion, their judgments were influenced most by the Y dimension, followed by the X, and then the Z. Therefore, under the rotation transformation the Z dimension was even further obscured. The S_s' statements regarding which dimension most affected their decisions was supported by the scaling analysis. For all three groups of S_s , the original Y dimension was present in the 3-dimensional, Euclidean space, for two of the groups, the X dimension was present as well, but, the Z dimension was never present. It therefore seems that the proportionality changes under the rotation transformation are the most likely cause of the lower consistency.

The exchangeability requirement for probability of de Finetti as interpreted by Wise (1971) has not been conclusively supported for personal probabilities; however, the model proposed by Wise (1970) appears to be an adequate representation of the way S_s generated subjective probabilities in this particular empirical task. Of course, the construction of the stimuli encouraged such a categorization, but the important point is that a mathematical formulation was possible which predicted subject probability estimates in a related task. In general, the results of this study definitely support the notion of an underlying cognitive structure which can adequately be described through the use of proper mathematical formulations.

References

- De Finetti, B. La prevision: Ses lois logiques, ses sources subjectives. *Ann. Inst. Henri Poincare*, 1937, 7, 1-68.
- Hoijer, B. On testing metric and nonmetric models of multidimensional similarity. 68th Report, Department of Psychology, University of Uppsala, s:t Larsgatan 2, S-75220, April 1969.
- Hyman, R., & Well, A. Judgments of similarity and spatial models. *Perception and Psychophysics*, 1967, 2(6), 233-248.
- Hyman, R., & Well, A. Perceptual separability and spatial models. *Perception and Psychophysics*, 1968, 3(3A), 161-165.
- Isaac, P. D. Linear regression, structural relations, and measurement error. *Psychological Bulletin*, 1970, 74, 213-218.
- Kruskal, J. B. Nonmetric multidimensional scaling. *Psychometrika*, 1964, 29, 1-27.
- Wise, J. A. Origins of subjective probability. *Acta Psychologica*, 1970, 34, 287-299.
- Wise, J. A. Symmetry, structure, and subjective probability. *Proceedings of the Third Research Conference on Subjective Probability, Utility, and Decision-making*. London, England, 1971, in press.

TELEVISION DISTORTION AND ATTITUDE CHANGE

A BACKLASH EFFECT

Michael Ozerkevich

Royal Military College of Canada

It was hypothesized that the effect of presenting video distortion during the televised description of an event by an individual would be for the viewers to rate the individual less favorably than another individual presenting his version of that event without the accompanying distortion. It was hypothesized that the distortion would be operating within the context of a classical conditioning paradigm. Results indicated that operant rather than classical procedures were likely to be operating since the finding that the effects of the introduction of the distortion was that subjects rated the individual in the distorted version as being more favorable than the individual in the undistorted version. This finding could not be explained satisfactorily within the classical paradigm.

A serious concern has existed among certain sectors of the public and among social scientists that the use of various methods of presentation of television content might exert profound effects on the attitudes and behaviors of viewing audiences. For example, an impressive body of research exists on the relationship between exposure to symbolic aggression and aggressive behavior. (See Feshbach & Singer, 1971 for an excellent summary). However, little attention has been directed toward the relationship between the technological factors of television transmission and their effects on viewers' perceptions of the content of the transmission. The present study represents the first in a planned series of studies designed to assess the effects of various video and audio distortions on the attitudes and values of viewing audiences in a laboratory setting.

Every transmitting television studio is technically capable of eliminating (and, indeed, encouraged to do so by its audience) audio and video distortions in the transmitted signal. Each transmitting television studio is also capable of *inducing* audio and video distortion in the transmitted signal at any point. The possibility exists, therefore, that the systematic introduction of distortions into the transmitted signal might have profound effects on the attitudes and behaviors of viewers (such as on the attitudes and voting behaviors of the audience during political campaigns).

Indeed, a number of investigators have already applied operant techniques to television program presentation in order to modify a wide

range of behaviors. Baer (1962) demonstrated that withdrawal of cartoons has a punishing effect on young children. Lindsley (1962) achieved the same punishing effects by using attenuation of both picture and sound instead of abrupt withdrawal. Greene & Hoats (1969) were able to demonstrate that video and audio distortion can function as negative reinforcers in escape-avoidance and punishment contingencies.

The present inquiry extended the application of television distortion by Greene & Hoats to the area of attitude change. It was hypothesized that the effect of presenting video distortion during the televised description of an event by an individual will be for viewers to rate him less favorably than another individual presenting his version of that event without the accompanying video distortion. It was theorized that the pairing of the video distortion with the televised version would constitute a classical conditioning paradigm.

Method

Subjects

Seventeen male cadets and enlisted men and three female secretaries, all serving at the Royal Military College of Canada, participated as subjects.

Procedure

A Sony one-inch video tape recorder and attached monitor were modified and set up according to the design described by Lindsley (1962). This design permitted the remote controlling of vertical, horizontal, contrast, and brightness distortions by the experimenter. Two film clips were edited from a commercial film (*The Eye of the Beholder*) and placed on video tape. Each clip contained a single narrative by a male actor describing another actor in the film. Each clip of narrative differed in its content, in the actor speaking and in the evaluation made of the third actor. Both clips were two minutes long.

All subjects were presented with one clip undistorted and with the other clip horizontally distorted (i.e., the picture rolled from bottom to top at a rate of approximately two frames per second). The order of the presentation of the clips and the clip which was distorted were counterbalanced. Ten subjects viewed clip number one distorted and clip number two undistorted. Clip number two was distorted and clip number one was undistorted for the remaining ten subjects. Subjects were tested in groups of four.

After each film clip was presented, subjects completed an 18-item version of Fiedler's (1967) LPC scale to describe the actor delivering the narrative. The version of the LPC employed followed Osgood's Semantic Differential (1952) and contained 18 bipolar adjective items.

Total scores were obtained for each subject's evaluation of the actor-narrator. A low score indicated that the actor had presented a favorable image to the subject. A high score indicated that the subject had perceived the actor unfavorably.

It was predicted that undistorted versions would produce lower scores than distorted versions. This prediction followed from the assumption that the distortion served as a noxious UCS and that the film clip served as the CS in a forward conditioning procedure. It was also assumed that the presenting of the UCS with the CS would produce a CR in the form of an unfavorable evaluation.

Results

Mean ratings on the LPC scale were obtained for both the distorted and the undistorted narrations as well as for each narrator under both conditions. The results of the comparisons of these means can be summarized as follows:

a. The distorted film clips produced significantly lower scores than did the undistorted film clips ($t = 2.50$, $df = 19$, $p < 0.05$ two-tailed and $p < 0.025$ one-tailed). The predicted higher favorability for the undistorted clips was not found. Instead, subjects perceived either actor in a significantly more favorable light when his narration was visually distorted than the other actor whose narration was not distorted.

b. When LPC scores were compared for each actor in both conditions the same preference in the opposite direction to that which was predicted was found. The first actor was perceived more favorable when his narration was distorted than when it was undistorted ($t = 1.23$, $df = 18$, $p < 0.20$ two-tailed and $p < 0.10$ one-tailed). The same higher degree of favorability for the distorted narration by the second actor was found ($t = 2.06$, $df = 18$, $p < 0.10$ two-tailed and $p < 0.05$ one-tailed).

The finding that subjects perceive narrators more favorably when their narration is visually distorted is difficult to explain in terms of the classical conditioning paradigm hypothesized to be operating here. If the effect produced in the present inquiry is a product of classical conditioning then the only possible explanation of the results is that the distortion introduced is, in fact, pleasing to subjects. This seems unlikely.

It is possible that subjects perceived the narrator in the distorted condition as being unfavorably presented through no fault of his own and, hence, rated him more favorably because of his "underdog" position. It is also possible that distorted versions represent a loss of information for the viewer and, in the absence of information, subjects tended to rate narrators on whom they had less information

more favorably than those on whom they had more information. Both possibilities represent operant conditioning paradigms and both can be examined experimentally.

The findings of the present inquiry indicate that the modification of viewers' attitudes through the use of distortions in the transmitted television signal might not be achieved through classical conditioning procedures. Rather, it may be that the procedures that are required are operant in nature. If this latter possibility is the case, then the individual who would modify attitudes through the use of various kinds of signal distortion would require a knowledge of the other reinforcers operating for each individual that would make the establishing of a contingency between the onset or offset of the signal distortion and a given attitude or behavior difficult or impossible.

The potential for attitude modification through the use of signal distortion in television broadcasting is tremendous and frightening if the distortion operates according to classical conditioning procedures. If, on the other hand, the distortion operates according to the principles of operant conditioning, then the broadcaster employing signal distortion to attempt to modify attitudes may find himself facing a "backlash" effect in which the attitudes change in the opposite direction to that which he had intended simply because other more powerful reinforcers were operating. An Orwellian future is more distant under the operant hypothesis.

References

- Baer, D. M. Laboratory control of thumbsucking through the withdrawal and re-presentation of positive reinforcement. *Journal of the Experimental Analysis of Behavior*, 1962, 5, 525-528.
- Feshbach, S., & Singer, R. D. *Television and Aggression*. San Francisco: Jossey-Bass, 1971.
- Fiedler, F. *A theory of leadership effectiveness*. Toronto: McGraw-Hill, 1963.
- Greene, R. J., & Hoats, D. L. Reinforcing capabilities of television distortion. *Journal of Applied Behavior Analysis*, 1969, 2, 139-141.
- Lindsley, O. R. A behavioral measure of television viewing. *Journal of Advertising Research*, 1962, 2, 1-12.
- Osgood, C. E. The nature and measurement of meaning. *Psychological Bulletin*, 1952, 49, 251-262.

DIVORCE AND FAMILY DISSOLUTION IN A
MILITARY ENVIRONMENT

John W. Williams

United States Air Force Academy

Previous research in the area of family dissolution shows that certain variables are highly correlated with divorce rates. Using UOR and sample survey data on USAF officers from 1958-1970 divorce rates were correlated with education, grade, rated/non-rated, flight specialty, component, command, source of commission, religion and Southeast Asia tour. Divorce rates of officer and the civilian population were also compared. This study indicated that divorce rates among Air Force officers are lower than is popularly believed.

Almost 4000 Americans break up their marriages every day of the year. In 1970 there were 715,000 divorces, involving almost one and a half million people. What is even more significant is the fact that the divorce rate has been increasing in a very dramatic way during the past ten years. Although I am sure your primary interest is in divorce and family dissolution in the military environment, I believe a brief examination of this phenomenon in the overall society is necessary in order to provide a theoretical orientation and framework.

To begin, let us look at data depicting the divorce rate per 1000 population for the years 1958 through 1970 (1958 was used because a curvilinear regression model on divorce data from 1950 through 1969 showed 1958 as being the low point for this 20-year period).

Examination of Table 1 reveals that the number of divorces increased every year but two. Decreases during these two years were slight; one-half of one percent in 1960 and two-tenths of one percent in 1962. The increase from 368,000 divorces in 1958 to 715,000 in 1970 was a 94 percent increase. We can see from Columns three and four that the rate per 1000 population also increased significantly during this period--from 2.1 in 1958 to 3.6 in 1970. This is more than a fifty percent increase, leading us to the conclusion that although the increase in population may be accounting for *some* of the divorce rate increase, it certainly doesn't account for all. Since 1966 the rate has been increasing more rapidly than before.

A statistical technique known as curvilinear regression analysis was used on this data in order to observe the trend of the line of best fit. The line of best fit shows that the divorce rate was decreasing until 1958 when an upward trend began which has shown no sign of abating.

TABLE 1
Divorce Rate per 1,000 Total Population 1958-1970*

Year	Number of Divorces	Rate per 1,000 Total Population	Percent Change in Rate
1958	368,000	2.1	
1959	395,000	2.2	+ 4.5
1960	393,000	2.3	+ 4.5
1961	414,000	2.3	0
1962	413,000	2.2	- 4.3
1963	428,000	2.3	+ 4.5
1964	450,000	2.4	+ 4.3
1965	479,000	2.5	+ 4.2
1966	499,000	2.5	0
1967	523,000	2.7	+ 8.0
1968	582,000	2.9	+ 7.4
1969	660,000	3.3	+14.0
1970	715,000	3.6	+10.0

*Divorce figures including reported annulments summarize monthly reports from 39 states and D.C. These areas contain over 80% of the population (rounded to the nearest 1,000).

Another way to observe the divorce trend is to look at the divorce rate per 1,000 marriages, as shown in Table 2.

TABLE 2

Divorce Rate per 1,000 Marriages

Year	Marriages	Divorces	Divorce Rate Per 1000 Marriages	Percentage Change in Rate
1958	1,451,000	368,000	253	---
1959	1,494,000	395,000	264	+ 4.0
1960	1,523,000	393,000	258	- 3.0
1961	1,548,000	414,000	367	+ 3.0
1962	1,577,000	413,000	262	- 2.0
1963	1,654,000	428,000	259	- 1.0
1964	1,725,000	450,000	261	+ 0.7
1965	1,800,000	479,000	266	+ 2.0
1966	1,857,000	499,000	268	+ 0.8
1967	1,927,000	523,000	271	+ 1.0
1968	2,059,000	582,000	283	+ 3.0
1969	2,146,000	660,000	301	+ 7.0
1970	2,179,000	715,000	319*	+ 6.0

*Provisional

Table 2 shows that there were 1,451,000 marriages in 1958 and 368,000 divorces. This gives a divorce rate per 1000 marriages of 253. This is generally the basis for the popular notion that the divorce rate is about one in four; however, from the 1970 data we can see that for every 1000 marriages that took place during this year 328 divorces occurred. This suggests that the divorce rate in the United States is probably closer to one in three than one in four, as popularly believed.

Regression analysis was again used and the same upward trend was observed. When this analysis was done, only data through 1969 were available and the model predicted that 333 divorces would be occurring for every 1000 marriages by 1975; however, we recently learned from the census bureau that the rate was already 328 in 1970, so it appears that the upward trend may be intensifying.

In order to better observe the divorce trend we also ran statistical analyses of such data as the divorce rate per 1000 married females 15 years of age and older, marriage and divorce number, rate and percent change, and marital status of the white and non-white population. Any way you look at it, there is no doubt but that divorce rates are increasing in a dramatic way.

Before dealing directly with data showing divorce among Air Force officers I would, very briefly, like to discuss a few theoretical aspects of this phenomenon. It will probably never be known what "causes" divorce. The best we can hope to do is specify conditions associated with divorce or point out significant interactions among several factors. When speaking of causes, one thing must be pointed out--the *legal* cause of the divorce is seldom the *actual* cause. Most divorces are awarded on the basis of mental cruelty, desertion, drunkenness, insanity, adultery, neglect to provide, and conviction of a felony. For example, 95% of the divorces in California in 1966 were filed on the one ground of "extreme cruelty." It cannot be said that adultery *causes* divorce because we know that in many cases it does not; however, those factors that previous researchers say are associated with divorce should be pointed out. As we go through these keep in mind your knowledge of the Air Force officer corps and how it relates to these variables.

Income appears to be related to divorce. All of the sociological literature very clearly points out that as income goes up, divorce rates go down. William Goode constructed a "proneness to divorce" table by income which showed that those who were most divorce prone were those in the lower income brackets. Where do Air Force officers stand in the income hierarchy? Contrary to popular belief Air Force officers have relatively high incomes. For example, compare the pay scales for officers and all civilians for 1969 as shown in Table 3.

In all cases the pay of Air Force officers is higher. This table does not include such fringe benefits as free medical care, BX, commissary, lower tax base, etc.

TABLE 3

Comparative Pay Scales, Civilian and Military , *1969

Population	Pay Scale		
	Median	Mean	% Making Over \$10,000.00
<u>Civilians</u>			
Year round full time workers	\$ 8,455.00	\$ 9,346.00	35.0
Professional, technical, and kindred workers	11,750.00	12,816.00	62.5
Managers, officials, and proprietors (except farm)	11,015.00	12,417.00	56.0
<u>Military</u>			
Air Force officers	\$13,400.00	\$15,598.59	88.1

*Military pay averages pay of all officers, both flying and non-flying.

A second variable that is related to divorce is education. Empirical studies conclusively point out that as education goes up divorce rates go down. Marital adjustment studies indicate that those with a high level of education have a higher level of marital adjustment than those with a low educational level. How does the education level of Air Force officers compare with that of civilians? Data on Table 4 show this comparison for 1969.

TABLE 4

Education Level of U. S. Civilians and Air Force Officers, 1969

Population	Education Level				
	Median Yrs. of School Completed	College Degree or More	5 Yrs. or More of College	Grad- uate Degree	Less Than 5 Yrs. of High School
Air Force Officers 153	16.5	83%	21%	18%	0.2%
U. S. Civilians 154	12.4	21%	10%	Unk	27.0%

Note that over 80% of Air Force officers have a college degree. Almost 20% hold graduate degrees. These figures point out that Air Force officers are highly educated.

A third variable related to divorce is *security*. Many sociological studies point out that economic insecurity is one of the factors giving rise to the great amount of divorce and family dissolution. Hollingshead mentions that economic security is one of the principal goals of the American middleclass family and that this is probably tied in with the fact that divorce is rare for this group. Those couples who have a good, stable income, money in the bank, little or no indebtedness, and a fair amount of life insurance are more likely to have fewer marital problems because of the security involved. Air Force officers should have strong feelings of security since they have a guaranteed income, free medical care for self and family, low-cost government insurance, an unequaled retirement plan, and a great amount of job security.

A fourth variable associated with divorce has to do with the kinship group. Most empirical studies show that the farther the physical and psychological distance from kin and in-laws the greater the chance for a successful marriage. The further you live away from your mother-in-law the greater your chance for marital happiness!

The very nature of service as an Air Force officer requires the family to live great distances away from in-laws, often in different countries. In-laws have very little opportunity to influence and disrupt the household.

A fifth variable associated with divorce is the visibility of the marriage. When both partners are well known and where the community can observe their behavior, there are greater restraints against social transgressions which may lead to divorce. This is tied in with community stigma. One of the barriers against divorce is community disapproval, and we note that villages and small towns have much lower divorce rates than urban areas. The marriage and family life situation for Air Force officers is generally highly visible. This is especially true for high-ranking officers, astronauts, and many officers in special assignments. Most Air Force bases and stations are very similar to small towns where many people are usually aware of what is going on in other families. The group we associate and live with has a great influence on our decisions and the visibility of the marriage has a beneficial effect on relations between husbands and wives.

Another variable associated with divorce has to do with occupations. Professional and managerial occupations, generally speaking, have high marital stability. Many researchers point out that jobs which provide a high level of intellectual or creative satisfaction, good income, and some degree of prestige create conditions most favorable to marital happiness. Air Force officers appear to meet these criteria very well.

In summary, variables correlated with marital happiness include income, education, security, occupation, mobility, visibility of the marriage, community stigma, propinquity of kin and in-laws, religious affiliation and separation.

On the other side of the coin are those variables which appear to be correlated with *higher* divorce rates. These include family separation, unfaithful behavior, conflicting religious beliefs, childlessness, mobility, and many others. The two that are most applicable to Air Force officers are separation and mobility. It is true that Air Force officers are highly mobile. They pack up and move many times over a 20 or more year career. What most people don't realize is that Americans in general are highly mobile and that packing up and moving is a way of life for many many civilian occupational groups.

Separation is a variable that appears to be correlated with higher divorce rates. Many researchers point out that the absence of the loved one and the anxieties about the welfare of family members subject marriages to far-reaching strain. Without doubt, family separation brings on some instability in the home life of the couple, especially if the separation is a lengthy one; however, it is my hypothesis that

short, periodic separations are functional for the marriage and that being away from our loved ones for short periods of time makes us appreciate them more.

After examining all of these variables associated with divorce it was my hypothesis that Air Force officers would have low divorce rates. They have high income, high education, a great amount of security, live far away from in-laws, and hold professional status in the occupational structure--and you will recall from what has been said before that all of these are positively correlated with marital happiness and success in marriage.

The data on Air Force officers were taken from UOR records maintained by the Air Force Human Resources Laboratory at Lackland AFB and from sample survey results conducted by the Data Services Division in the office of the DCS/P at the Pentagon.

Table 5 shows Officer Force Marital Status data from 1960 through 1965. Divorced status was not recorded until 1963. You will note that the percentage who were in divorced status during this period was about one percent.

Table 6 shows data from 1966 through 1970. Again we note that the number of officers divorced is about one percent. This is in the face of rapidly increasing numbers and percentages in American society. Since UOR data are updated every six months and officers are *required* to show changes in their marital status, there is no reason to believe that many officers are divorcing and remarrying and not showing up on the divorce data. In fact, we recently conducted a longitudinal computer run and discovered that only about 500 officers divorce each year and that about 90% of them eventually remarry--but very few remarry within 6 months of their divorce. In order to get a better picture of divorce and family dissolution among Air Force officers, we made a careful and detailed examination of available data. First of all, we looked at divorce by educational status from 1963 through 1970. Those officers holding the doctorate, including MD, PhDs, and others, had the lowest incidence of divorce. Those having less than a college degree, the highest. This supports the finding of previous researchers that marital happiness is related to higher education. Chi-square analysis of the data proved to be significant at the .001 level.

The next variable we were concerned with was Rated/Non-rated. For all years the percentage of divorced rated officers was higher than non-rated officers. We carried this one step further and looked at rated and non-rated officers broken down by grade. The major differences occurred at the rank of captain and major. There was quite a large spread for captains for all years. We found only one general and very few colonels in divorced status over the period of this investigation.

TABLE 5

Officer Force Marital Status 1960-1965

Marital Status	Year											
	1960		1961		1962		1963		1964		1965	
	N	%	N	%	N	%	N	%	N	%	N	%
Married	*	86.3	*	87.7	*	86.6	113,627	85.3	112,789	83.8	108,673	83.4
Single	*	13.6	*	12.4	*	13.5	17,051	12.8	19,381	14.4	19,913	15.3
Divorced	*	*	*	*	*	*	1,465	1.1	1,615	1.2	1,436	1.1
Legally Separated	*	*	*	*	*	*	400	0.3	135	0.1	130	0.1
Widowed	*	*	*	*	*	*	400	0.3	270	0.2	130	0.1
Annulled	*	*	*	*	*	*	*	*	*	*	*	*
Unknown	*	*	*	*	*	*	266	0.2	403	0.3	261	0.2
Total	*	99.9	*	100.1	*	100.1	133,209	100.0	134,593	100.0	130,543	100.0

* Unknown

Officer Force Marital Status 1966-1970

* Unknown

Table 7 depicts divorce means for different flight specialties. Of all the flight specialties, flight nurses consistently have the highest percentage of divorced officers. We don't know yet if these officers are entering active duty in a divorced status or if they divorce while on active duty and remain in service. We suspect that the former is true.

TABLE 7

Mean Divorced by Flight Specialty

Pilot	1.2	Flight Nurse	2.1
Navigator	1.4	Flight Medical Officer	1.2
Observer	1.6	Astronaut	0.0
Flight Surgeon	0.7	Misc. - No Rating	0.9

Concerning divorce and component, we found that reserve officers have slightly higher rates than regular officers. We were also interested in divorce rates in the different commands. If your experience has been the same as mine, you have probably heard throughout your career about the high divorce rates in SAC. We found no evidence to support such a proposition. Available data point out that the SAC divorce rate is no higher than for other commands and that the rate has shown no increase since 1963. It may be that the rate was higher in the early fifties, but there is no empirical evidence to verify this. I searched diligently for such evidence and found none; however, the rate for TAC *does* show an upward trend, as does the rate for MAC.

We also examined data on divorce by source of commission, shown in Table 8. Those officers who were graduates of OCS consistently showed up as having the highest divorce rate. Second were those who were commissioned through Aviation Cadets. Neither of these sources required a college degree for commissioning. The lowest rate was found among Air Force Academy graduates. This may be partially explained by the fact that the Academy graduated its first class in 1959; however, in 1970 only 21 of the 4,490 Air Force Academy graduates on active duty were divorced. Fifteen of these were pilots. Additionally, we know that if divorce is going to occur it will occur fairly soon after marriage, so many of these graduates *have indeed had time* to get divorced.

The data indicate that those officers who were Catholic or Jewish had the lowest divorce rates while those with no religion or no preference had the highest rate. This supports previous research findings which indicate that being attached to some religious organization is correlated with lower divorce rates.

TABLE 8

Five Year Divorce Means for Sources of Commission

Source of Commission	Mean Divorce Rate
Officers Candidate School	1.5
Aviation Cadets	1.4
Direct Appointment, Civilian	1.3
United States Naval Academy	1.3
Warrant Officer	1.2
Direct Appointment, Military	1.0
United States Military Academy	0.8
Officers Training School	0.8
Reserve Officer Training Corps	0.8
United States Air Force Academy	0.4

We also looked at data by race and sex and found negligible differences between Blacks and Whites. This is quite different from the situation in American society where we find that Blacks have much higher divorce rates than Whites. An hypothesis established as a result of this finding is that when income, education, housing, and life style are approximately the same, *many* statistical differences between Blacks and Whites will vanish. We did find that female officers have consistently higher divorce rates than male officers.

We found further that those officers who have had a Southeast Asia tour have somewhat higher divorce rates than those who had not.

We were able to compare divorced status of Air Force officers with that of males in the civilian population for the years for which data were available. For all years the percentage for civilians was higher. What is even more significant is the fact that civilian percentages show an upward trend while those for Air Force officers remain constant. We also made this comparison for females and found that although female officers have higher rates, the gap is closing. We were also able to compare the number of divorced men per 1000 married men and again found that rates for officers are lower.

The two major conclusions reached in this study were these:

1. There has been a dramatic increase both in numbers and percentages of divorce over the past ten years in American society.
2. Divorce rates for Air Force officers are relatively small, show no upward trend, and reflect favorably on the cohesive nature of the military family.

Low divorce rates among Air Force officers can be partly explained by reference to several sociological concepts. Primary among these is that of integration, i.e., the societal integration that takes place through shared norms, values, and beliefs. The sharing of common prescriptions for conduct, beliefs, and valuations promotes group solidarity. Marital cohesiveness is analogous to group cohesiveness and can be defined accordingly. Group cohesiveness is the total field of forces which act on members to remain in the group and to hold strong feelings of loyalty to the group. Thus, the strength of the marital relationship would be a direct function of the attractions within and the barriers around the marriage. A group which highly internalizes the norms, values, and beliefs of the group is likely to show strong solidarity. In certain military environments the sharing of a common culture and adherence to common norms, values and beliefs is profound. The Air Force officer corps is a homogeneous, stable group in which a common set of standards and goals is shared by practically all members. These norms, values, standards, and goals are generally shared by wives.

Although the officer force to some degree cuts across all social classes, there is a strong feeling of commonality of kind. In fact, young officers are socialized by both their peers and their superiors to direct their loyalty toward the group and toward the mission. Selfless devotion to country and to the Air Force is strongly encouraged. They are also encouraged to put away feelings of superiority and desire for individual recognition and work toward success of the squadron, group, or wing. This is integrative and leads to solidarity. This loyalty to the country, to the Air Force, to the unit, more than likely carries over into loyalty to the wife.

The relatively high income and the fringe benefits received by the Air Force officer and his family as well as the outstanding retirement system promote strong feelings of security--a concept highly correlated with low divorce rates. The fact that the husband and wife share mutual involvements, patterns of interest, and participate mutually in many external activities contributes to marital cohesiveness.

Previous research has suggested that when the couple is separated from both kin groups the marriage has a better chance of success. Air Force couples fit this criterion very well. The very nature of military service is such that the partners usually live great distances from the kin groups. In-laws have very little opportunity to interfere and wives find it difficult to go home to mother--especially when living thousands

of miles away. The fact that the couple must work out their own problems free from interference and advice from kin-groups is functional for the marriage.

According to sociological theory, the great amount of mobility associated with military family life should lead to high rates of disorganization, including divorce. When civilian families move to a new location, they generally move into a completely new life style--a new job, new neighbors with different values, beliefs and ways of doing things--a different sub-culture, etc.

For the Military family this is not generally true. Most Air Force bases are similar and the lifestyle changes little from one to another. There is much less all-round family disruption due to the fact that there are many service agencies set up to assist the military family when they move. These internal supports act as inducements for the couple to remain in the group and as factors to ease the friction and disruption caused by the high mobility rate. When separation occurs through TDY or remote PCS moves, the Air Force acts somewhat like an extended family and offers comfort and assistance when the husband is gone. It is probable that this sense of group cohesiveness carries over into marital cohesiveness.

In conclusion, I would like to state again that divorce rates among Air Force officers appear to be rather low. In any given year only about one percent of the officer force is divorced. The great increases found in the proportion divorced in the overall population do not hold for Air Force officers.

It is not within the scope of this paper to attempt to explain the upward trend in the divorce rate in the U. S. population; however, any explanation would have to include such considerations as changing divorce laws, the greater affluence of this country, lessening social stigma for divorced persons and the fact that divorce is in the process of becoming institutionalized. This study points out that divorce rates in American society began an upward trend in 1958 that shows no sign of leveling off.

STRATEGIES FOR CONDUCTING MISSION-ORIENTED RESEARCH

IN MILITARY ORGANIZATIONS

Arthur B. Sweney

Wichita State University, Wichita, Kansas

The Center for Human Appraisal and Communication Research obtained a grant from the Air Force Office of Scientific Research to conduct an action-research program in the 381st Strategic Missile Wing to study and effect change in areas relating to retention, quality of life, and leadership effectiveness. A consultative relationship was established with the wing command staff. Various programs were instituted to generate interest in the research process and to insure meaningful feedback concerning the results as they became available. The programs instituted to implement this included: educational credit for officers conducting research, special briefings, periodic project director reports, a project newspaper for circulation to all members of the unit and a system of external collaboration with other investigators interested in behavioral research in the Armed Forces.

Since the enactment of Congressional Bill 91-441 and the rider commonly referred to as the "Mansfield Amendment," research in the Armed Forces of the United States has taken on new dimensions which need not be categorized as bad. "Unit-connected, mission-oriented, and command-approved research" provides a challenge to the conscientious investigator which cannot be hidden behind scientific jargon or idiosyncratic research interest. It must be conceded and emphasized that all research cannot and should not fall within these constraints, but there are, never-the-less, no indications yet that these strictures need render research conducted in the military either weak or meaningless.

Strategies

Applied vs Pure Research

In the field of psychology and many other sciences there has been a tendency to dichotomize research efforts into the classes of "pure" and "applied." This distinction is viewed by the author to be both spurious and detrimental to the development of meaningful and actionable inquiries. "Pure science" has often become identified with the controlled observation of trivial laboratory induced variations. By extracting the essence of life, which is complexity, from human behavior, it is possible to obtain simple enough sequences to manipulate and measure, and these frequently are the only studies able to qualify as "pure science."

Applied science, on the other hand, has too frequently been identified with ungeneralizable measurement. The studies of unique samples measured under unique conditions has provided science with a plethora of fact but a dirth of theory. Applied science has tended to engage the dragon of unknown with a frontal assault rather than nibbling at his tail in the "pure science" tradition.

Method-centered vs Mission-centered Research

The layman and even some scientists have some very romantical concepts concerning the dynamisms underlying scientific development. Concepts such as "discovery" and invention have taken on mystical colorations which isolate them from the sequences and processes which constitute the fabric of science. By most people science is perceived as being a grand mobilization of efforts against problems and challenges, thus engaging the dragon head first. More frequently, however, science has not been "mission-oriented" but "method-oriented" and has ignored the problems of non-understanding and have become method-centered. One experiment followed another in orderly fashion with each giving greater clarity to those that preceded.

Method-centered research has received much of its impetus from the development of instruments and research technologies. The microscope, telescope, mass spectrograph, vacuum tube, transistor, accelerator, wind tunnel, analytic balance, galvanometer, qualitative analysis, artificial bacterial cultures, material testing machines, oscilloscopes are just a few examples of instruments in the physical sciences which have generated method-centered research.

The mission or the problem is frequently the focus of applied research. Headlong assault is made upon the unknown with the recognized ignorance as the primary tool. Tychociner (1950), in his study of *Zetetics* (the science of research), points out that areas of scientific inquiry can be computer generated from known technology and existing instruments, but are hard to organize and difficult to predict when launched with a mission-orientation. Most funding agencies and the general population seem to accept an agenda that recognizes "mission-oriented" research as having top priority.

Examples of successful mission-oriented research are few, but never-the-less noteworthy. The Oak Ridge Project, The Rand Corporation, the efforts of the Salk and Sabine vaccines are historic examples. Whether the problems of pollution or military personnel retention will generate meaningful mission-oriented research remains to be seen. Its success may depend upon the degree to which method-oriented research has already developed a technology and scientific base which can be integrated and successfully applied.

Testing Models vs Describing Samples

One of the primary questions which lies behind the development of a research strategy is the question of purpose and focus. Is the purpose to expand the theoretical foundation for viewing behavior, human or otherwise? Or is the focus the description and analysis of a specific organization, individual or sample? This differential emphasis is part of the hidden agenda contributed by the investigator. Scientists most frequently are testing implicit models or hypotheses. Statisticians, technocrats, and publication hungry academicians are usually satisfied to describe samples.

Although this distinction is easy to outline abstractly, it is much less easy to distinguish in actuality. Principles are often developed from an exhaustive effort to "understand" or explain a single organism. The concept of friction may very well have evolved from the study of the behavior of a single block of wood on an inclined plane. This was a theoretical not a descriptive study since the data was eventually translated into the theoretical and generalizable properties exhibited.

The Research Study vs the Research Program

The layman and many respectable scientists view "discovery" as being the result of a single crucial experiment or study. Because of the publication lag and the latency in reporting this often seems to be reflected by the published articles. In most cases, however, the single article is the culmination of a program of research in methodology, instrumentation, and foundation principles which has been integrated together through the self-correcting discipline of repeated studies and replicated outcomes.

The scientific community can ill afford the Quixotic although creative thrusts of the impulsive investigator. Isolated positive or negative outcomes of a single study can furnish false leads or signal the end for what would otherwise be fruitful lines of inquiry. If the investigator is insufficiently motivated to replicate his own studies he should not be too surprised that other scientists do not show significant interest.

Treatment Research vs Investigatory Research

Since the time of Bacon's "*Novum Organum*" the scientific method has been primarily mobilized as an investigatory tool. Few of us would question the legitimacy of this as a primary focus. In the last thirty years, however, there has been a growing awareness that research actually changes the phenomena which it studies. Evidence of this fact has been accumulating from both the physical and social sciences. The Hawthorne studies were "contaminated" by changes in the workers' behaviors which were induced by the fact that they became aware that

they were being studied. The electrical engineer has found it impossible to monitor an electrical circuit without in some way changing the properties of that circuit no matter how loosely he couples his measurement instrument to it.

Cattell in his theories of personality has concluded that man's behaviors become more systematic as he understands and accepts as "true" those theories and principles which have been developed to systematically explain it. Cattell suggests that the number of factors in a personality questionnaire may reflect to some degree the complexity of the subjects' theories concerning personality. George Kelly (1955) in his Psychological Personal Constructs makes similar kinds of observations. He posits that each person is his own behavioral scientist who tests hypotheses concerning his own behavior and the behavior of others. As his theories are confirmed or infirmed he organizes his behavior into a more meaningful pattern. In this way the feedback from research as well as a person's own answers to questions may profoundly affect his future behavior.

Prescriptive vs Descriptive Research

Normally the scientist conducts research from an objective point of view. He hesitates to make the kinds of value judgements which are often needed if advice is to be given. He normally describes what exists and makes no prescriptions. He feels that his duty has been properly executed if he has properly exposed problems and their antecedent conditions. He usually does not feel called upon to personally marshall the forces necessary to change those things which he has found to be detrimental to individuals or society. Mission-oriented research tends to reverse some of these trends. If the "mission" is to be accomplished changes must occur in prescribed directions.

Prescriptive research varies from descriptive research in some other important ways. The investigatory phase of this kind of research must be more selective in its scope. In order to be prescriptive it must focus upon actionable variables. Other kinds of science are free to study any facet of a problem. Prescriptive research is primarily concerned with those variables that can be manipulated and changed. For this reason, demographic or ontogenetic variables are of little interest to the prescriptive researcher since they are not actionable, i.e., they cannot be manipulated or changed to obtain the desired results.

The prescriptions which arise out of research take numerous forms which in turn seem to suggest varying probabilities of success. In the behavioral sciences the three major options for recommendations seem to be: changing people, changing the peoples' perceptions of the systems, and changing the systems. The former is often fraught with unpredictable side effects although some behavior modification experiments have been very successfully executed, particularly with

mental defective or infrahuman subjects. Changing systems and individual's perceptions frequently provide more fruitful alternatives.

Summary

Table 1 illustrates some of the tactics, procedures, and programs utilized in AFOSR Project #2001 to implement these various strategic considerations. No effort has been made to be inclusive in this paper.

TABLE 1: SAMPLE OF STRATEGIES AND TACTICS FOR THE CONDUCT OF RESEARCH
ON AFOSR Project #2001

Strategies	Tactics	Procedures - Programs
I. Applied vs Pure research	A. Seek generalizable results B. Naturalistic C. Manipulative Research	A. 1. Inferential statistics 2. Stratified samples 3. Multivariate analysis B. 1. Study unit as is 2. Obtain unobtrusive measurements. C. 1. Institute subtle treatments 2. Monitor results
II. Method Centered vs Mission Centered Research	A. Utilize familiar instruments and methods B. Focus on mission C. Divergent research approaches Focus upon single missions	A. 1. RPM and related instruments 2. Correlational techniques B. 1. Retention 2. Unit Effectiveness 3. Quality of life C. 1. Multidisciplinary approaches 2. Research Teams
III. Model Testing vs Sample Description	A. Test- Response to Power Model (RPM) B. Test Model of man as a communication C. Provide sample descriptions	A. 1. people perception 2. Superior-Subordinate influence 3. Rating Bias B. 1. Communication Diary 2. Communication Tests C. 1. Unit Description 2. Rank Description
IV. Research Study vs Research Programs	A. Continue "program" studying RPM B. Integrate "programs" instituted by others C. Conduct isolated "studies" in novel related areas	A. 1. Compare Air Force results to industry 2. monitor changes B. 1. Adapt Herzberg's instruments 2. Use Cattell's 16 PF C. 1. Social conformity 2. Human Resource Accounting
V. Treatment Research vs Investigatory Research	A. Repeated measurement B. Inform subjects of objectives C. Inform subjects of results, group or individual	A. 1. Replicate in same unit. 2. Replicate in Different unit B. 1. Briefings-Preliminary 2. Informing Commanders and Officers 3. Give purpose of testing C. 1. Reports 2. Entrails articles 3. Post Research 4. Individual feedback sessions. 5. participative research
VI. Prescriptive vs Descriptive	A. Study actionable variables B. Look for implied policy changes C. Prescribe alternative system and procedures for a consideration and evaluation.	A. 1. Study roles 2. Aberrations in power and communication systems B. 1. Learn policy, 2. Study effects of policies C. 1. Couch recommendations in actionable terms 2. Recognize influence of larger systems not studied which dictate ultimate decisions

FACTORS AFFECTING GROUP DECISION-MAKING

Hal W. Hendrick

United States Air Force Academy

and

Charles R. Holloman

Virginia Polytechnic Institute and State University

Reviews the results of a series of studies in which various structural and process variables were manipulated to determine their effect on the quality of group decisions. Group size, type of problem, relative esteem of leader, and decision-making procedures were found to affect decision accuracy. Implications of these results for actual decision-making groups were discussed.

During the past four years we have been conducting a series of programmatic studies on group decision-making. This research has been concerned with those structural and process variables which can be controlled easily in real life situations as well as in the laboratory. This paper summarizes our more important findings to date.

Individual versus Group Decisions on Factual and Non-factual Tasks

In our first project we investigated the effectiveness of individuals as compared to small groups in solving two basic kinds of problems (Holloman & Hendrick, 1970). The first of these are problems in which the solution is dependent on knowledge of specific factual material. The second kind of problem is one in which the problem solution is not dependent on a specific body of facts.

Tasks were chosen which would be similar in format of participation and response but differ in the degree to which outside background information could facilitate individual performance.

The factual problem was the NASA Decision-making Exercise. This task is described as follows:

"You are in a space crew originally scheduled to rendezvous with the mother ship on the lighted surface of the moon. Due to mechanical difficulties, however, your ship was forced to land at a spot some 200 miles from the rendezvous point. During reentry and landing, much of the equipment aboard was damaged and, since survival depends on reaching the mother ship, the most critical items available must be

chosen for the 200-mile trip. Below are listed the 15 items left intact and undamaged after landing. Your task is to rank them in order in terms of their importance in allowing your crew to reach the rendezvous point. Place the number 1 by the most important item, the number 2 by the second most important, and so on through number 15, the least important."

The nonfactual task involved the movie *Twelve Angry Men* which depicts the jury deliberations at the completion of a murder trial. The initial vote of the jurors is an 11 to 1 vote for guilty. The movie was stopped at this point and individuals were told that, during the remainder of the movie, the 11 jurors who initially voted guilty changed their votes, one by one, resulting in a 12 to 0 vote of not guilty. Each juror exemplified a distinct personality, and his arguments and behaviors prior to the initial vote suggest a degree of possible behavioral rigidity or flexibility. Individuals were asked to predict the order of changeover of the 11 jurors from voting guilty to voting not guilty.

Method

The subjects were 80 cadets enrolled in an advanced leadership course at the USAF Academy. Ss were divided into groups of five to seven each. For both tasks, after each individual had finished his private prediction they were asked to reach a single group prediction using the rules of consensual decision-making:

1. Avoid arguing for your individual judgments. Approach the task on the basis of logic.
2. Avoid changing your mind *only* in order to reach agreement and avoid conflict. Support only solutions with which you are able to agree somewhat, at least.
3. Avoid "conflict-reducing" techniques such as majority vote, averaging, or trading in reaching decisions.
4. View differences of opinion as helpful rather than as a hindrance in decision-making.
5. View your initial agreement as suspect.

Results

Individual and group predictions were compared to the correct solutions. Net differences were referred to as error scores. The results for this study are summarized in Table 1. From examination of the table it may be noted that, for both tasks, the mean group consensual error score was smaller than the mean of the average

Table 1
Summary of Results

Error Score		Gain or Loss	
Average of group members	Most accurate group members Group consensus	Over average error score	Over most accurate individual
Non-factual Task			
N = 80	14	14	14
$\bar{X} = 20.6$	12.1	6.9	-1.6
SD = 2.1	2.4		
Factual Task			
N = 80	14	14	14
$\bar{X} = 38.4$	27.1	9.4	-1.9
SD = 2.9	3.6		

individual error scores for each group. Also, for both tasks the mean of the most accurate group member error scores was found to be slightly smaller than the group consensual mean error score.

In order to determine if these mean differences for both tasks were significant, t tests for correlated data were computed. For the non-factual task, the observed t for the difference between the mean of the average individual error scores and mean group error scores was found to be 7.95. Entering the t table with $df = 13$, this difference was found to be significant at the .001 level. For the factual task, the observed t was 4.1. With $df = 13$ this was significant at the .005 level. The observed t for the difference between the mean group consensual error score and the most accurate group member mean error score was 1.6 for the non-factual task and .7 for the factual task. With $df = 13$, both of these were found not significant at the .10 level.

It was concluded that:

1. Group consensual decision making was more accurate than averaging individual decisions for both tasks.
2. On the average, the decision accuracy of even the most accurate group member was not significantly better than the group consensual decision.
3. The magnitude of the difference between group consensual decisions and the average of individual decisions was greater on the non-factual task than on the factual task.

Influence of High Status Members on Group Decisions

In another study Dr. Holloman and I were concerned with the extent to which group members perceived as having high status influenced group decisions (Holloman & Hendrick, 1972b). In particular we wondered if, given a status hierarchy, there was a considerable gap in the hierarchy between the leader and the other group members, would the leader tend to exert more influence than when there was not a large gap between the leader and the other group members.

Method

The subjects were 82 upper class cadets enrolled in an advanced leadership course at the USAF Academy. For the first 14 sessions of the course, Ss participated in group discussions of case studies and reading materials and took part in various classroom exercises involving various dimensions of leadership behavior. During this period, each section conducted a peer ranking of its members using the criteria of overall quality of participation and degree of interpersonal influence exercised in the classroom. This ranking was openly discussed in the sections and was consensually validated. Because

of its sensitivity to the influence component (Bass, 1960) this ordinal ranking was recognized as reflecting the leadership hierarchy of the section.

The Ss were then selectively assigned to groups under condition 1 or condition 2 according to their ranking in the status system of their section. In the assignment of Ss to groups, each group had as its highest status member either the first, second, or third ranked member of the section. The mean rank of group leaders under condition 1 was 1.7. For groups under condition 2, the mean rank of group leaders was 1.9. Group leaders under condition 1 were separated from persons immediately below them in the status hierarchy of the section by at least three ordinal positions. The membership of groups under condition 2 reflected a continuum of the status hierarchy of their sections. Groups varied from five to seven subjects in size. Ss then participated in the *Twelve Angry Men* exercise following the procedure outlined above for the first study.

Results

The private prediction of each S was compared to the correct solution. The sum of the absolute differences between the rankings was determined. The difference between each S's private prediction and the correct solution was identified as his error score. The difference between each S's private prediction and his group's consensual solution was identified as his influence score. The error score for each group was also determined. For each group the following data were computed: (a) error score and influence score of the group leader, (b) error score and influence score of the most accurate member of the group, and (c) error score of the group.

Table 2 reflects the results of comparisons between performance data of the groups under the two treatment conditions.

The homogeneity of the groups under the two treatment conditions was determined through comparisons of the error scores of the group leaders, the most accurate members of the groups, and the groups' consensual solutions. No significant differences were found in these between-treatment comparisons. Group leaders of the discontinuous-status groups were slightly more accurate on the average than were the leaders of the continuous-status groups; however, this difference lacked significance at the .05 level. In two instances, the leaders of discontinuous-status groups were also the most accurate members of their groups. Group leaders in the discontinuous-status groups were more influential in causing their groups to accept their private solutions as the group's solution ($t = 2.83, p < .02$). Because of the small number of groups under each treatment and the resultant few degrees of freedom, and the relatively large differences in the standard deviations, the Mann-Whitney U-test was also used to compare these sample means. The resulting value of U was 5 ($p < .02$).

Table 2

Summary of Comparisons: Between Groups

	Discontinuous Status Groups		Continuous Status Groups		Diff.	df	t
	N	\bar{X}	SD	N	\bar{X}	SD	
Error scores of group leaders	6	15.0	5.47	7	18.6	3.15	3.6 11 1.41
Error scores of most accurate member	6	11.0	2.09	7	13.1	2.27	2.1 11 1.64
Group error scores	6	12.7	3.44	7	14.1	2.34	1.4 11 .84
Influence scores of group leaders	6	6.0	4.50	7	15.7	7.54	9.7 11 2.83*

* $p < .02$

Leaders of the discontinuous-status groups were more successful in having their private predictions accepted by their group as the group's consensual prediction than were the leaders of the continuous-status groups. The error score and the influence score of the group leaders were compared with the respective data of the group's most accurate member. Using a one-tailed test of significance, the most accurate members in groups, under both conditions, were significantly more accurate in their private predictions than were the leaders of their groups. In the discontinuous-status groups the resulting t was 2.40 ($p < .05$) and in the continuous-status groups the resulting t was 4.50 ($p < .005$). Under treatment one, group leaders were more influential than were the most accurate members of their respective groups ($t = 2.64$, $p < .05$). Under treatment two, no differences were found in the influence scores of the group leaders and the most accurate members. The ability to influence other group members was greater for the leaders of the discontinuous-status groups than for other members of the groups. In the continuous-status groups, the ability of the group leaders to influence the group was no greater than the ability of the most accurate member.

Group Decisions as a Function of the Decision-making Process

In a third study, we investigated the adequacy of group decisions as a function of the various common procedures that can be used in arriving at decisions (Holloman & Hendrick, 1972a).

This study was designed to determine the adequacy of six techniques of group decision-making when (a) groups are homogeneous with respect to ability, (b) groups have a history of interaction, (c) the task is non-factual in nature and involves a wide range of possible solutions, and (d) time limits are uniform. These techniques may be differentiated along a continuum according to the pattern and quantity of social interaction required and/or permitted. The six techniques are as follows: (a) averaged decision of individual members, (b) decision by Chosen Leader, (c) Minority Control, (d) Majority Vote, (e) Consensus, and (f) Consensus after Majority Vote. It was hypothesized that the adequacy of the group decision process would be positively related to the amount of social interaction involved in the process.

Method

The Ss were 137 upper class cadets enrolled in 12 sections of an advanced leadership course at the USAF Academy. For the first ten weeks of the semester, Ss participated in the ongoing learning activities of their class section. These activities consisted of discussions of case studies and related reading materials and participation in classroom exercises involving various dimensions of leadership behavior. During this period each section member became aware of each other member, the resources he brought to the class section, and his method and pattern of participation. Although these 12 sections weretaught by five different instructors, there was considerable uniformity

of course content and teaching methodologies between sections. During the 20th class session, Ss in each section were randomly divided into groups of six each. These groups remained intact through the end of the experimental exercise. From the 20th through the 25th class sessions, the groups participated in a variety of problem-solving and leadership exercises.

Beginning with the 26th lesson, Ss participated in the Twelve Angry Men exercise. For Technique 1, the private decisions of the members of each group were statistically averaged and the resulting decision was viewed as the group's decision. These private decisions were collected before any interaction occurred, and the intended use of the collected private decisions was not announced. The groups were then told that instead of participating directly in the making of a group decision, they would instead choose a leader who would make the decision for the group. At this point Ss were asked to indicate their choice of a member of their respective group who would make the group's decision. Printed forms were passed out to the Ss for indicating their choices. After these forms were collected, Ss were asked to think of a second person whom they would like to choose to help the leader in making the group's decision. The private decision of the chosen leader was taken as the decision resulting from Technique Number 2. The person chosen as group leader and the member chosen as his assistant were then asked to compare their private decisions and resolve any differences into a single, mutually acceptable decision. The resulting decision was viewed as Technique Number 3, the Minority Control Technique. In working toward a single decision between them, the two chosen leaders were permitted to interact only with each other.

Those groups designated to employ the Majority Vote Technique in reaching a group decision were instructed to vote separately on the order of change of each juror. Discussion was permitted only as required to reach a simple majority vote. After this process was completed, the group decision was collected. These groups were then instructed to continue in their process but that the objective was now a consensual decision.

Groups using Technique Number 5, Consensual Decision, were given the same instructions as were groups using Technique Number 6. These groups worked toward a consensual decision without prior use of the majority vote technique.

Results

The private decision of each S was compared to the correct solution. The sum of the absolute differences between each S's private prediction and the correct solution was referred to as his error score. Error scores were also computed for each group's final decision. Table 3 summarizes the mean error scores of the group decisions by decision technique.

Table 3
Summary of Results: Group Error Scores

	1	2	3	4	5	6
	Averaged Decisions	Chosen Leader Decision	Minority Control	Majority Vote	Consensus	Consensus after Majority Vote
Number of Groups	23	23	23	12	11	12
\bar{X}	20.52	18.48	18.22	16.42	12.55	9.67
SD	3.82	6.76	4.26	3.00	3.44	2.23

In Technique Number 1, the private decisions of each group member were averaged to produce the group decision. Interaction in using this technique was limited to the extent that each individual decision was weighted equally with that of each other member. As was hypothesized, this technique resulted in the least accurate decision. At the other end of the continuum, the Consensus after Majority Vote Technique, which requires more interaction than any of the other techniques, resulted in the most accurate group decisions. Also, for each of the four intermediate techniques, our data show that decision accuracy was positively related to the quantity of interaction.

Differences between mean error scores resulting from the six decision techniques were tested for significance by analysis of variance. All groups were involved in Techniques 1, 2, and 3. Twelve groups used the Majority Vote Technique and upon completion also used the Consensus Technique. Eleven groups used the Consensus Technique.

The overall F for the Consensus after Majority Vote groups of 17.2 was significant beyond the .01 level. For the Consensus groups, the overall F was 6.04, and was significant at the .05 level. In order to determine which of the comparisons were significant, a Newman-Keuls *a posteriori* test of ordered pairs of means, was conducted. Technique Number 6 was found significantly more accurate than Techniques 4, 3, 2, or 1 ($p < .01$). The Majority Vote Technique was also more accurate than Technique 1 ($p < .05$) but did not differ significantly from Techniques 3 and 2. The Consensus Technique was significantly more accurate than Techniques 3 and 2 ($p < .05$) and Technique 1 ($p < .01$). Comparison between the Consensual Decision Technique and the Majority Vote Technique and between the Consensual Decision Technique and Consensus after Majority Vote was made through use of the t test for independent means. The consensus procedure was significantly more accurate than the Majority Vote procedure ($t = 2.87, p < .05$).

In summary, the data reveal no significant differences between the first three procedures used. While there are numerical differences in the error scores of the final decisions resulting from these procedures, these differences lack significance. As shown in Table 3, there is a trend toward increased accuracy as the kind of decision-making technique used by the groups was changed to permit more interaction. Only the last three procedures, however, permitted the sort of interaction which fully meets the definitional requirements of the interaction process. The data show that majority voting is superior to the Averaged Decision Technique but is not significantly more accurate than the decisions made by chosen leaders or those resulting from minority control. The decisions resulting from Techniques 5 and 6 were, respectively, superior to the other techniques investigated, all of which required less group interaction.

Decisions as a Function of Group Size

In a recently published study (Holloman & Hendrick, 1971), effect of group size on decision accuracy was reported.

Method

The Ss were senior and junior cadets enrolled in 18 sections of various sociology and social psychology courses at the USAF Academy. Ages of Ss ranged from 19 to 24 with a modal age of 22. For the purpose of participating in the experimental task, each section was randomly divided into various combinations of differently sized groups depending upon the number of students in the section, which ranged from 16 to 22. Data were collected from 269 Ss in 8 groups of 3, 19 groups of 6, 1 group of 5, 5 groups of 9, 5 groups of 12, and 5 groups of 15. The one group of 5 originally had 6 members; one member was absent at the time of the exercise. For the first 12 weeks of the semester, Ss participated in the ongoing learning activities of their course and class section. During the 30th class session, Ss were randomly divided within sections into groups of size 3, 6, 9, 12, or 15. The groups remained intact and worked as discrete groups through the end of the experimental exercise. The purpose of this early division was to allow the group members to become more fully acquainted with the other members of their group, and his beliefs, abilities, and his pattern of participation in the group. From the 31st through the 35th class hours, the groups participated in a uniform variety of behavioral exercises related to the sociology and the social psychology disciplines. Beginning with the 36th session, Ss participated in the Twelve Angry Men exercise previously described.

Results

The effect of size upon group interaction and the accuracy of the resulting consensual decisions was tested for significance through analysis of variance. Table 4 summarizes the analysis of the error scores of the group consensual decisions. The resulting F of 12.35 is significant beyond the .01 level. In order to determine which of the comparisons differed significantly, a Newman-Keuls *a posteriori* test of ordered pairs of means was conducted. As shown, groups of 3 were significantly less accurate ($p < .01$) than all other sized groups. Groups of 9 were less accurate ($p < .05$) than groups of 6 but did not differ from groups of 12 or groups of 15.

The increases in decision accuracy may be attributed to the effects of social interaction since holding the size of the groups constant isolated the effects of statistical considerations. Between groups of different sizes, interaction had a differentiated effect on the quality of the consensual decisions. Groups of 3 were least accurate of all other groups tested. In working toward a single decision with which they could all somewhat agree, the members of groups of 3 necessarily had to depend upon the information and ideas

TABLE 4

Summary of ANOV--Error Scores of Group Consensual Decisions

Source	SS	df	MS	F
Between Groups	647.22	4	161.81	12.35**
Within Groups	3,459.01	264	13.10	
Total	4,106.23	268		

NEWMAN-KEULS

GP Size	3	9	12	15	6
\bar{X}	18.25	14.20	13.22	13.04	12.49
		**	**	**	**
					*

* $p < .05$ ** $p < .01$

available to them. The lack of an obvious answer made it necessary that the groups actually choose from the available information and ideas. In this process of choosing from the large number of available alternatives, the error-correcting properties of group interaction were less present and effective. Wrong sets tended to persist, which resulted in less accurate final decisions.

Groups of size 6 produced the most accurate decisions, although their accuracy did not differ significantly from groups of 12 and groups of 15. Actual observations of the groups during the experimental exercise offer some clues to understanding these outcomes. First, groups of 6 required only minimal effort to organize themselves to work on the task. Little hierarchical differentiation was observed and questions of leadership and power did not arise. Groups of 9 attempted to operate in the same manner as did groups of 6 but were less effective in assimilating the contributions of all the members into the final decision product. In groups of 9, with each member feeling the freedom to participate, the fixed time limits had the effect of giving less time to each participant. Not only does the time available per member for overt communication diminish as the size of the group increases, but the pattern of communication also varies. In observations of groups of 12 and groups of 15, there appeared to be an undue concern with "getting organized so we can get the job done." These groups attempted to make decisions about choosing a discussion leader and about the operational meaning of consensus; however, no final decisions about getting organized were made. What actually happened was a factionalization of the larger group into two subgroups. As the members learned that some individuals began actively to make suggestions about the problem task, they began to defer to this small minority; and the suggestions and opinions of the active participants were more frequently incorporated into the final decision. Thus one subgroup became **actively involved** in the task and was responsible for the final decision product. The second subgroup became noticeably passive with the result that participation became very unequal. This observation is consistent with conclusions by Kelly & Thibaut (1954) that increasing the size of problem-solving groups increases the restraints against participation with the result that an increasingly large proportion of the group is discouraged from making overt contributions. The active subgroups within the groups of 12 and 15 tended to interact more like the groups of 6 did, with the less active members counting for less than their equal share of the total volume of interaction.

References

- Bass, B. M. *Leadership, psychology, and organizational behavior*. New York: Harper, 1960.
- Holloman, C. R., & Hendrick, H. W. Individual versus group effectiveness in solving factual and non-factual problems. *Proceedings, 78th Annual Convention, APA*, 1970, 673-674.
- Holloman, C. R., & Hendrick, H. W. Problem solving in different sized groups. *Personnel Psychology*, 1971, 24(3), 489-500.
- Holloman, C. R., & Hendrick, H. W. Adequacy of group decisions as a function of the decision-making process. *Academy of Management Journal*, in press.
- Holloman, C. R., & Hendrick, H. W. Effects of status and individual ability on group problem solving. *Decision Sciences*, in press.
- Kelley, H. H., & Thibaut, J. W. Experimental studies in group problem-solving process. In G. Lindzey (Ed.), *Handbook of Social Psychology*, Vol 2. Reading, Mass: Addison-Wesley, 1954, 735-785.

Training

Minority Groups

CHAIRMAN: Major Hal W. Hendrick

IMAGINATION AS A FLIGHT SIMULATOR

Dirk C. Prather

United States Air Force Academy

Twenty-three Ss were randomly placed in one of two groups. All Ss were student pilots and minimally experienced in the landing of the T-37 aircraft, the independent variable. The experimental group (E) listened to four twelve and a half minute tape recordings which prompted their mental practice of landing the T-37 aircraft. The control group (C) did not receive this practice. All Ss were rated by their instructor pilots on procedures and ability to land the aircraft on the mission that followed the last mental practice session. Group E's ratings on both procedures and ability to land were significantly higher ($p < .05$) than the ratings of Group C. It was concluded that the use of mental practice may be an effective adjunct to any training program which normally depends on costly actual practice of the skill being learned.

With the rising cost of simulation devices it is important to evaluate other devices and techniques that may be able to improve performance in a perceptual-motor skill. Mental practice of a skill is where the S attempts to vividly imagine the perceptual-motor actions involved in practicing the skill. Davis and Wallis (1961) have found that regular mental practice is superior to irregular actual practice in motor skill learning. Twining (1949) found no significant differences between actual and mental practice on basketball foul shooting. Shick (1970) was able to improve a volleyball skill through mental practice. Blurton (1969) used behavior therapy with imagery to significantly improve field goal shooting in practice, but found no significant differences in actual game situations. It appears that mental imagery can, in many cases, improve performance of a perceptual-motor skill.

The author, in an unpublished study, attempted to improve strafing in student fighter pilots through mental practice. He found that mental practice of this skill did improve actual strafing scores over those that did not use the mental practice technique. Due to loss of control over the experimental subjects, statistical analysis was impossible.

Corbin (1967) found that some previous experience with the skill is necessary for mental practice to be effective. It was decided that landing an aircraft by low experienced student pilots would be a skill

in which the Ss had minimal experience, yet is a highly complex perceptual-motor skill of the type that would be important to investigate. If this skill could be improved by mental practice, then it would strongly suggest that many less complex human skills may also be improved by this technique. This experiment was pointed toward improving performance in flight training in the United States Air Force. The Ss had some experience in landing an aircraft, but very little in landing the particular aircraft that was the independent variable. Due to the problems the author encountered in the control of the Ss in his pilot study, it was decided to use tape recordings as a prompt to the mental practice. This allowed for an exact timing of the student mental practice and a more precise control of his mental imagery. By weighing the student time and the cost of the apparatus, the cost effectiveness of such a program could be compared to more sophisticated methods of simulation.

The question proposed in this research was whether four highly prompted mental practice sessions, of approximately twelve and a half minutes each, could improve the student pilot's performance on landing an aircraft.

Method

Subjects

The subjects were 23 randomly selected student pilots in the undergraduate T-37 pilot training program at Williams Air Force Base. Thirteen were in the experimental group (E) and ten were randomly placed in the control group (C). All Ss were low experienced student pilots with approximately 20 hours in the T-41 trainer and 4 hours in the T-37.

Apparatus

The experimental sessions for the E Ss were conducted in the learning center at Williams AFB. This center has typical student learning carrels for individual instruction through media presentation. The E Ss sat in a cockpit procedures trainer of the T-37 aircraft. This cockpit mock-up was configured similar to the actual aircraft through photographs. The only movable items in this mock-up were the throttles and the control stick. The instructions and stimulus information were played through earphones over a dial access tape recording.

Procedure

The E Ss had observed and attempted the experimental task, that of landing the T-37 aircraft; but this experience was at a low level consisting of approximately 7 previous landings. The E Ss were

instructed to go to the learning center after they had completed the fourth, fifth, sixth, and seventh mission in the flying training syllabus and listen to a tape recording while sitting in the cockpit mock-up.

The tapes were designed to give instruction in the landing pattern. The E Ss were told to imagine the situations as vividly as possible and to perform the same motor actions and eye movements that they would if they were in the actual landing pattern. In the first few imagined landing sequences the E Ss were given complete instructions as to the airspeeds, throttle settings, pitch attitudes, bank required, etc. In the later imagined patterns the cues were withdrawn until in the last few sequences the tapes merely stated "You are on base" or "You are on final." To vary the sequences slightly, error analysis, go-arounds, touch-and-go, and final full-stop landings were all covered in this experimental training. The running time for each tape, in order, was 11:50, 15:10, 11:20, and 10:45.

The C Ss were not given any of the above experimental training. These C Ss received the normal training that past student pilots have received, which included some media presentations in the learning center.

After the eighth actual flying mission both the E and C Ss were rated by their own instructor pilots on their performance as to technique and procedures in the landing pattern on that particular mission. This was a relative rating of the student's performance on several areas in the landing pattern. The instructor pilots did not know which students were in which group. Several instructor pilots had a student in each group to rate.

Results

The Ss' instructor pilots filled out a one to seven rating scale on techniques and procedures for the following phases of the landing pattern: initial to pitch, pitch to 180, 180 to final, final to flare, flare to touchdown, and go-around. The ratings for these phases of the landing pattern were averaged for each of the techniques and procedures area to give a more meaningful, stable rating. The procedures area was defined as how well the student knew what to do and the techniques area was defined as how well he actually did the landing task. The rating was relative in that the instructor was told to rate the S in relation to all the other students he had instructed on that particular mission.

The results were analyzed by means of the Mann-Whitney U test. On procedures, the E group had a mean rating of 4.53 and the C group 4.26 ($U = 35.3$, $p < .05$, two-tailed). On techniques, the E group had a mean rating of 4.21 and the C group 3.89 ($U = 38.0$, $p < .05$, two-tailed).

Discussion

From the results of this experiment it appears that mental practice combined with actual practice is more effective than just actual practice when learning a perceptual-motor skill. The tape recorded presentation, using withdrawal of prompts to help control the mental imagery, is probably more effective than just letting the student imagine the skill without structure. Further structure was added to the mental practice by having the S sit in the cockpit mock-up of the aircraft he was flying. With the extra practice gained by using prompts, it might be expected that the mental practice would improve the procedures of the S; but the finding that the actual performance was improved through transfer of the skill practiced in the mental imagery sessions is very significant.

All E Ss filled out a critique on the program. Without exception they felt the mental practice helped them to perform better while flying. Most of the E Ss stated that they did not have any problem in vividly imagining the situations called for by the tape recordings.

Because the independent variable involved in this experiment is a highly complex perceptual-motor skill, the results can probably be extended to include many areas of skill learning. The use of mental practice may be an effective, low-cost adjunct to any training program which normally depends upon costly actual practice of the skill being learned.

References

- Blurton, R. R. Effects of group behavior therapy imagery on basketball performance. *Dissertation Abstracts*, 1969, 29(9-B), 3476-3477.
- Corbin, C. B. The effects of covert rehearsal on the development of a complex motor skill. *Journal of Genetic Psychology*, 1967, 76, 143-150.
- Davis, E. C., & Wallis, E. L. *Toward better teaching in physical education*. Englewood Cliffs, New Jersey: Prentice-Hall, 1961.
- Shick, J. Effects of mental practice on selected volleyball skills for college women. *Research Quarterly*, 1970, 41, 88-94.
- Twining, W. E. Mental practice and physical practice in the learning of a motor skill. *Research Quarterly*, 1949, 20, 432-435.

NAVY FIGHTER PILOTS' RESPONSES TO PAPER-AND-PENCIL
SIMULATIONS FOR AIR-TO-AIR COMBAT SITUATIONS

Richard S. Elster and William H. Githens

U. S. Naval Postgraduate School, Monterey, California

This paper addresses the information used by Navy fighter pilots in making decisions in air-to-air combat. The primary decision investigated was that of whether or not to engage the enemy. Simulated air-to-air combat situations were described on questionnaires mailed to members of Navy VF squadrons. The pilots' responses were analyzed as a function of range, bearing, and other variables woven into the situations described.

The research described here was stimulated by discussions with some students at the Naval Postgraduate School who had flown combat missions over North Vietnam. These discussions led to a feeling that different pilots might make rather different decisions when facing the same enemy air-threat. As a first step in the exploration of the variables used by pilots in making decisions in air-to-air combat situations, a mail-out survey was conducted of Navy fighter pilots to obtain their rating of the relative importances of a number of variables possibly involved in air-to-air combat decisions.

The results of this initial survey, plus the results of interviews of fighter pilots at the Postgraduate School, led the investigators to choose six variables for investigation. These variables were: the maximum amount of fuel that can be used in engaging the enemy (i.e., fuel above bingo), enemy range, enemy bearing, enemy heading, number of enemy aircraft, and the rules of engagement (eyeball or missiles free). The objective of this research was to capture the pilots' decision making policies regarding air-to-air combat decisions and focused on studying the aforementioned six situational variables.

Method

Instruments

An "air-to-air threat evaluation" questionnaire was constructed in order to obtain fighter pilots' decisions with regard to specific situational conditions in which the six variables under investigation are systematically varied. Each of the six variables was specified at one of two values for each of the tactical situations. In all, then 2^5 or 64 combinations of the variables were considered by each fighter pilot. The variables and the two values of each variable that were studied are displayed in Table 1.

TABLE 1

Tactical Situation Variables and Their Values

	Fuel (Above Bingo)	Enemy's Range	Enemy's Bearing	Enemy's Heading	No. of Enemy Aircraft	Rules of Engagement
Values	1000 lbs	20NM	315°	045°	2	Eyeball (EB) ²
Assigned	2500 lbs	Eyeball (EB) ¹	135°	225°	6	Missiles Free (MF) ³

¹. Within 4 NM.

². Positive visual identification required before firing.

³. Interpreted as meaning the pilot may fire without positive visual identification

As a final step in describing the tactical situations to be considered by the pilots, the following background scenario was used for all tactical situations:

"You are the flight leader of a section of F-4s armed with two Sparrows and two Sidewinders. Assume for this exercise that the aircrafts' weapon systems are up in every respect. You are providing TARCAP for a division of A-7s who have just completed a strike and are egressing from the target area. You are feet-dry over North Vietnam (20NM to the coast). The AAA and SAM defenses in the immediate area are light to moderate. You have limited GCI facilities operating for you and the enemy has excellent ground radar control.

The enemy aircraft are assessed to be MIG-21s at 15,000 ft. and 500 kts. You are 10,000 ft. and 450 kts. heading for your carrier (360° relative).

The weather in the area is clear and 15+ visibility. There are several flights of attack aircraft still feet-dry, exact position unknown.

The MIGs have recently demonstrated an air-to-air missile capability."

Figure 1 shows the way the tactical situations were presented to the respondents.

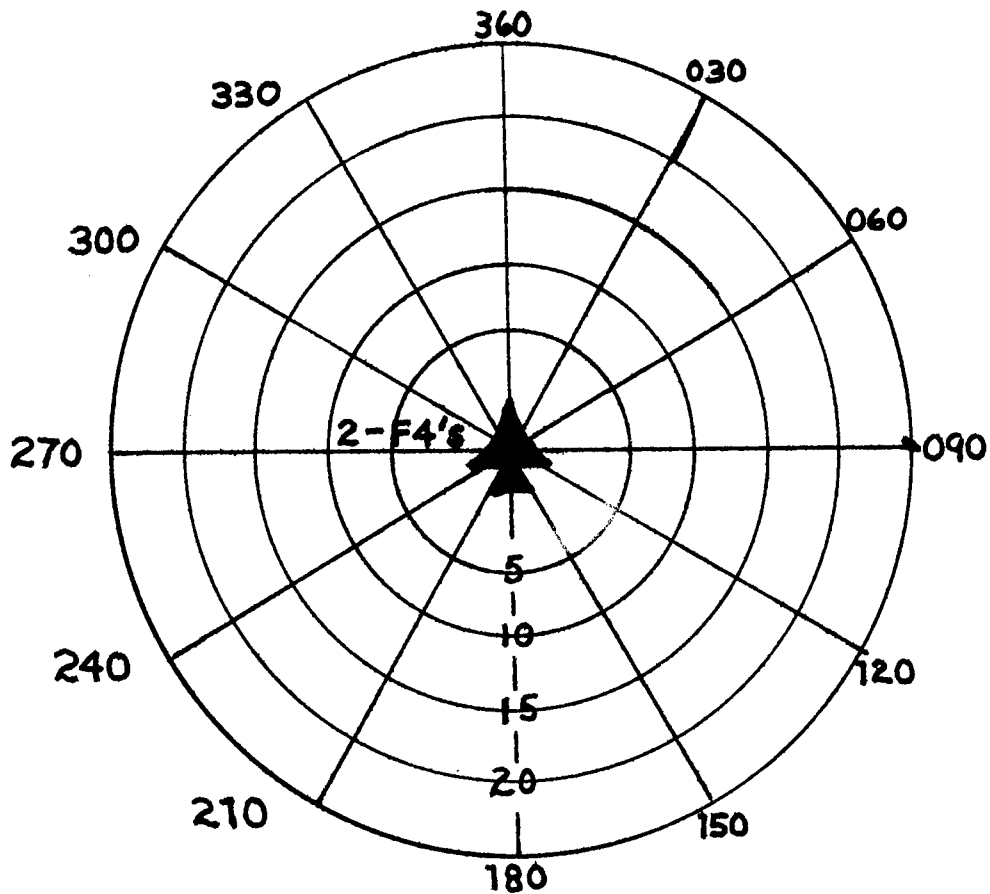


Fig. 1. Sample tactical data presentation.

Two major questions were asked of each pilot in conjunction with each of the 64 tactical situations he considered. These questions are shown in Figure 2.

- A. In this tactical situation (check one answer);
1. I'd have no choice; there would be an engagement. _____
 2. I'd have a choice on whether to engage or not, and I'd engage. _____
 3. I wouldn't engage. _____
- B. Indicate what aircraft losses you would predict, if there were an engagement.
- Enemy Losses _____ (no. of aircraft)
- Friendly Losses _____ (no. of aircraft)

Fig. 2. Questions asked with each tactical display.

The responses to the items in Figure 2 were then used as dependent variable data in the analysis conducted.

Subjects

The questionnaire containing the 64 tactical situations was mailed to the commanding officers of a number of Navy fighter squadrons with the request that he have each of his pilots complete a copy of the questionnaire. Thirty-six of the returns were complete and usable. Table 2 contains a description of the responding pilots in terms of their ranks, total flight hours, and number of combat missions.

TABLE 2

Description of the Sample of Fighter Pilots Who Responded¹

		LT(jg)	LT	LCDR	CDR
Number Responding		1	21	10	4
Total Number of Flight Hours	<u>Mean</u>	450.0	1394.3	2675.5	3825.0
	<u>S.D.</u>	0.0	438.54	553.69	466.60
Total Number of Combat Missions	<u>Mean</u>	0	96.59	163.4	201.8
	<u>S.D.</u>	0	74.66	72.17	106.80

- ¹ Pilots from 14 different VF squadrons responded to the questionnaire.

Results

As a first step in the analyses of the data, intercorrelations were computed among the variables of interest using the data from all 36 pilots. A part of the obtained correlation matrix is shown in Table 3.

TABLE 3

Correlation of Selected Variables with the
Tactical Decision Data from 36 Pilots¹

Variable	Correlations
Amount of Fuel	-.17
Enemy's Range	.42
Enemy's Bearing	
Danger Angle ²	.29
Enemy's Heading	
Number of Enemy	.12
Rules of Engagement (Eye/Mis)	-.08
Predicted % Enemy Killed	-.43
Predicted % Friendly Killed	-.23

¹ The three decisions were:

- 1) I'd have no choice; there would be an engagement.
- 2) I'd have a choice of whether to engage or not, and I'd engage.
- 3) I wouldn't engage

² Enemy bearing and headings were combined into "danger angle" using the following equation:

$$\text{Danger} = \text{Sign of sin (Bearing X [Bearing - (Heading + 180) Modulo 360]}$$

The data in Table 3 show that the two variables most related to the tactical decision were the enemy's range and the percentage of the enemy aircraft predicted as being killed. Of considerably more interest, however, is to see what is found when using multiple regression methods in attempting to capture the tactical decision-making policies of the pilots. The results of such a multiple regression analysis are shown in Table 4.

TABLE 4

Multiple Correlations and Standard Partial Regression Coefficients
Obtained When Predicting Tactical Decision Selection¹

Predictor	Respondents			
	LT(N=21)	LCDR(N=10)	CDR(N=4)	Aggregate (N=35)
Fuel	-.22	-.11	-.07	-.17
Danger	.29	.37	.19	.29
Eye/Mis	-.09	-.02	-.02	-.08
Range	.43	.38	.51	.42
No. of Enemy	.18	.15	.08	.12
Multiple R	.58	.56	.55	.56

¹ Since the predictors were established in such a way that they were uncorrelated, the entries in this table are identical to the predictor-criterion correlations.

The data in Table 4 show that the enemy's range was always the best single predictor of the respondents' tactical decisions and it should be noted that there are some intriguing hints of differences in decision-making policies among the three ranks represented, i.e., the Commanders, Lt. Commanders, and Lieutenants. Danger, a variable computed from enemy bearing and heading, was always the second best predictor.

In order to determine whether or not we could improve on the prediction of the tactical decisions made by the pilots, three biographical variables were added to the prediction equation. The results of this multiple regression analysis are contained in Table 5.

Table 5 also shows the results obtained from two other multiple regression analyses using the predicted number of enemy or friendly aircraft "killed" as criteria. These analyses show that the best predictor of the number of enemy predicted killed is the amount of fuel the (friendly) pilot has. More surprising, however, is the finding that the best predictor of the predicted number of friendly killed is the (friendly) pilot's rank, rather than one of the characteristics of the tactical situations.

TABLE 5

Multiple Correlations and Standard Partial Regression Coefficients
For Three Criteria--Data from Total Sample of 35 Pilots¹

Predictor Predictor Pre	<u>Dependent Variable</u>		
	Tactical Decision Tactical Decision	Predicted No. of Enemy Killed	Predicted No. of Friendly Killed
Fuel	-.17	.34	-.02
Range	.42	-.02	-.18
No. of Enemy	.12	.02	.25
Eye/Mis	-.08	.23	-.05
Danger	.29	-.29	-.02
Rank	-.02	-.01	.31
Hours	-.01	.04	-.13
No. of Missions	-.02	.07	.04
Multiple R	.57	.51	.38

¹ The one LT(jg) was deleted from the sample due to his lack of combat experience.

Discussion

The authors wish to mention some limitations and warnings that should be noted. First, only two values of each of the independent variables, e.g., fuel, range, number of enemy aircraft, rules of engagement, heading, and bearing, were presented to the pilots. The chosen values, and the particular combinations of the values, were judged to be reasonable by fighter pilots aiding with the development of the questionnaire, but it must be recognized that the entire range of any particular variable was not presented.

The second limitation that must be mentioned is that the multiple correlations presented here are very likely inflated. That is, they

have not been crossvalidated by using the responses of a second set of fighter pilots to determine the stability of the multiple regression equations that have been developed.

Lastly, we don't know the stability over time of an individual pilot's responses to the tactical situations contained in the questionnaire. One would hope that a trained pilot's responses to the same simulated tactical situation would be quite stable, but as yet the data necessary for investigating this issue are not available.

The multiple correlations we have obtained all have relatively low values. Assuming that the methodology of multiple regression can capture any existing policies, the lack of a high R might be explained by randomness in the decision-making. An alternative explanation would be that there is perhaps little randomness in actual combat decisions, but the determining variables have not all been included in our study.

We feel the possible implications of the findings presented here are most unsettling. If additional studies continue to show that fighter pilots' decisions are hard to account for by means of the variable describing the tactical situations, and that rank is the best predictor of the number of friendly losses expected, the results will have dramatic training and selection implications.

TRANSFER AND STRESS: A COMPARISON OF STUDENT PERFORMANCES RESULTING
FROM TRIAL-AND-ERROR AND PROMPTING AND FEEDBACK INSTRUCTIONAL
TECHNIQUES IN THE LEARNING OF A PERCEPTUAL SKILL

Gene A. Berry, Dirk C. Prather, and John M. Bermudez

United States Air Force Academy

Forty male students at the United States Air Force Academy were trained on a range estimation task. The Ss were randomly assigned to either Group RFB, that were given feedback by their actual range plus a verbal reinforcer if they were within a given range, or Group SFB, that were given feedback by their actual range plus a 60-volt electric shock if they were outside the given range. The learning curves showed no differences in the groups during training. After training to asymptote there were no significant differences on the transfer variable, but Group SFB's performance was superior under stress ($p < .05$). The results were discussed in reference to the possibility of training resistance to stress.

In several of the modern views of pedagogy it is suggested that to increase the efficiency of learning a student should not be allowed to make errors. Programmed instruction and the systems approach to learning both tend to take this position. Proponents of this viewpoint have stated that it is detrimental to the learning situation if the student has to unlearn an incorrect response. This position suggests that highly structured prompts should be used to make it nearly impossible for the student to commit errors. Other views of learning hold that it is necessary for the student to receive feedback on his performance in order to learn the correct response. In the highly prompted learning situation the necessity of feedback is lessened to the point that it is almost absent or becomes intrinsic to the situation. This lack of feedback may have an effect on the efficiency of learning, performance under stress, and the transfer of training.

Prather (1971) found that trial-and-error (T&E) learning of a perceptual skill was equal to highly prompted training on efficiency. The highly prompted technique had only a slight advantage after one or two trials but soon lost this advantage. When transfer of training and stress was the variable, T&E trained subjects performed significantly better than those trained by the highly prompted techniques. Prather & Berry (1970) extended the earlier results to another population. They trained both groups to asymptote and were able to plot the learning curves of both groups. Except for a nonsignificant advantage

to the highly prompted group on the first two trials, the two learning curves were almost identical. The significantly better performance by the T&E group under the transfer situation was confirmed.

Berry, Prather, & Jones (1971) took a further look at the effects of prompting in the learning of a perceptual skill. They trained one group using a combination of prompting and T&E techniques. The first three learning trials of this group were highly prompted, and then the following training trials were conducted under T&E conditions. The other group was trained under only T&E conditions. The learning curves paralleled the earlier study and were almost identical. Although both groups were trained to asymptote, the T&E group's performance was significantly better under the transfer conditions. There were no significant differences under stress. This research suggests that if a program is designed to transfer the learned skills and abilities to a new stimulus situation, prompting, even early in training, has a deleterious effect on performance.

The current experiment is a continuation of the studies cited above and was pointed toward flight training in the United States Air Force, as the task involved was one that is similar to those skills that a pilot must exhibit. This investigation was designed to determine whether feedback is desirable even under highly prompted conditions. Training was continued over a large number of trials to the point where each learning curve had virtually reached its asymptote. One group was given feedback as to their performance after they had been prompted by a highly structured cue. The other group was trained by a T&E, with feedback, technique. This procedure allowed (a) comparisons of whether prompting or feedback is necessary to learning, (b) an analysis of the learning curves over a large number of trials, (c) a comparison of performance under stressful conditions.

Method

Forty male students at the USAF Academy were randomly assigned to one of two groups. Every S had passed a stringent physical exam, which included eye and depth perception tests, within the 12 months preceding the experiment. Some Ss had minor corrections of their eyes to 20/20 vision. All these Ss wore their glasses during the trials.

The task was not a simple S+ or S- discrimination but one along a continuum. Each S was trained to select a discrete point at which he perceived a target to be a preselected distance from him. This skill is much like the ones a pilot must exhibit. The S was required to estimate the point at which a target that was closing toward him at 800 ft/sec was at a range of 2,000 ft. This was a simulation of strafing. The training stimulus was a black square filmed on 16mm. black-and-white film against a plain white backdrop. The camera was moved toward the simulated target at a speed to approximate 800 ft/sec closure rate over 8,000 ft. of range. Each run was filmed individually. Some were started at a shorter distance, and on others the camera was

operated for varying increments of time before the run was started. These two actions were taken to vary the number of seconds to the correct solution so that Ss could not merely estimate by means of elapsed time. The transfer target was a photograph of a MIG-21 aircraft, which was three times as large as the training target.

The sequence of the film was 26 runs on the training target and 3 runs on the transfer target. Approximately 5 sec. of black film was inserted between each trial to give E time to record the stopwatch readings and to give feedback when appropriate.

The Ss were seen individually in a classroom. They were trained by one of two methods: T&E or by a combination of highly cued and T&E techniques (CUE). Each S was seated 20 ft. from the screen and held an electric trigger button in his right hand. The T&E Ss were required to press the trigger button when they estimated that the target was at the correct range. On the odd-numbered trials, E gave S his range as feedback by means of a verbal statement, e.g., "2400 ft." The CUE Ss, on the odd-numbered trials, were prompted by a light, located just below the target image on the screen, that E illuminated when the target was at 2800 ft. of range. These Ss were required to press the trigger when they estimated that they were 2000 ft. from the target. Feedback was given as to their performance. This allowed the CUE Ss to be prompted, yet make the same motor response and receive the same feedback as the T&E Ss. The even-numbered trials were test trials. All Ss were required to press the trigger when they estimated they were 2000 ft. from the target, and they did not receive any feedback or cue light on these test trials.

On the 26th trial, the Ss were placed under a stressful condition, a performance-contingent electric shock. The film was stopped before this trial, and the electrodes were attached to the S's wrists. He was told that he would receive a mild, slightly painful shock if he was more than 400 ft. off from the 2000 ft. desired open-fire range. The S was not shocked, only the threat was used. The film was again stopped and the electrodes removed.

Trials 27, 28, and 29 were the transfer trials. The S was shown a picture of the MIG-21 and told its length and the size relationship of the new target to the old. He was told to try to estimate a range of 2000 ft. on this new target. No cue light or feedback was given on these trials.

Results

Learning curves for the training groups using the CUE and T&E techniques are portrayed in Figure 1. The absolute errors are indicated for each test trial throughout the training period. A low score indicates better performance than a high score.

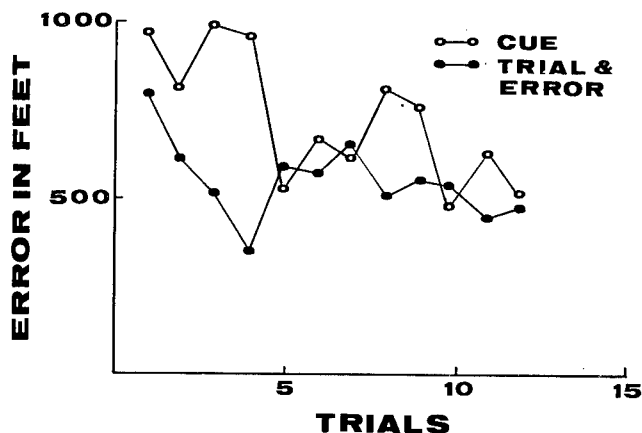


Fig. 1. Learning curves for the training groups using the CUE and T&E techniques.

The learning curves reflect that the T&E group performed better during the training period ($t = 3.06$, $p < .01$). This advantage was particularly pronounced during early test trials.

For the transfer trials the two groups had almost equal results. The average errors were 1051 feet for the T&E group and 1020 feet for the CUE.

The introduction of stress during the final training test trial had the expected effect of decreasing performance. The CUE group had an average error of 570 feet and the T&E group 520 feet. Although both groups suffered a performance decrement, there was no significant difference between the two methods in the stressed training situation.

Discussion

The most interesting finding of the present experiment was that performance was significantly better during acquisition without the use of cues. It appeared from the learning curves that there was some confusion early in the learning phase that might have been caused by the prompt. The curves tended to closely approximate each other late in the learning phase but were quite diverse early in learning. This finding was contrary to the previously cited literature, which indicated that the prompted group enjoyed a small but insignificant advantage after the first two or three learning trials. Possibly the CUE group relied too heavily on the prompt and did not receive as much benefit from the feedback as the T&E group. The extreme amount of variability in the CUE group's learning curve would support this view.

The finding that the groups did not differ on the transfer and stress variables indicates that the presence of feedback may be necessary to enhance these skills. The cited literature found significant differences in transfer of the learned skill in favor of the T&E method over prompting. In these studies the CUE groups did not receive feedback as to their performance during learning. This had the effect of depressing performance during learning. In the current experiment feedback was added to the CUE group's conditions. The addition of this feedback appeared to negate any advantage of strictly T&E learning.

Looking at the past studies and the findings of the current experiment, it appears that the extra expense of adding cueing or prompting devices to learning situations may not be justified. If there is a choice of whether to give the learner feedback or prompts, the evidence strongly supports the decision to use only feedback methods.

References

- Berry, G. A., Prather, D. C., & Jones, G. L. Effect of prompting and feedback on the learning of a perceptual skill. *Proceedings of the 79th Annual Convention of the American Psychological Association*, 1971, 6(2), 589-590.
- Prather, D. C. The efficiency of trial-and-error versus errorless learning of a perceptual-motor skill and performance under transfer and stress. *American Journal of Psychology*, 1971, 84(3), 377-386.
- Prather, D. C., & Berry, G. A. Comparison of trial-and-error versus highly prompted learning of a perceptual skill. *Proceedings of the 78th Annual Convention of the American Psychological Association*, 1970, 5(2), 677-678.

AN INVESTIGATION OF POSSIBLE TEST BIAS

IN THE NAVY BASIC TEST BATTERY

Patricia J. Thomas and Edmund D. Thomas

Naval Personnel and Training Research Laboratory

An analysis of data from 104,683 white and 2,067 black Navy Class "A" school students showed that while the Navy Basic Test Battery does not consistently underpredict the performance of blacks, the test validities for the black students were extremely low at some schools.

A great deal of research effort recently has been devoted to the study of test bias. Selection instruments used by colleges, industry, and government are being scrutinized to determine whether standards developed with a predominately white population are reasonably predictive of the performance of black (or other minority) populations. In general, test bias results from inappropriately applying performance estimate equations developed on the basis of a majority sample to minority groups. Consistent underprediction of the criterion scores of minority members is referred to as negative bias. Conversely, when the performance of the minority group is overpredicted it is referred to as positive bias.

A review of the relevant literature generally supports the conclusion that negative bias is not common. Cleary (1968) found no evidence of negative bias, although some positive bias was found in her investigation of the Scholastic Aptitude Test as a predictor of grades at three colleges. O'Leary, Farr, & Bartlett (1970) conducted seven studies of predictor-criterion relationships in job situations. They concluded that test bias did exist in the majority of comparisons between blacks and whites but that it did not necessarily discriminate against the blacks. Guinn, Tupes, & Alley (1970), working with an Air Force enlisted population, investigated differences in validities for various cultural groups. Regarding the question of racial differences, they found that the performance of blacks in technical schools was generally overpredicted, i.e., black students earned lower grades than would be expected from their test scores.

Typically, the Navy has not had enough blacks in most Class "A" schools to investigate whether its classification test battery, the Basic Test Battery (BTB) is discriminatory. While the absolute number of Negro enlisted men has not risen substantially over the past few years, the number of blacks assigned to schools has almost doubled. This has been due to a conscious effort on the part of the Navy to counter the criticism that blacks are too often assigned to the Steward rating or given other domestic-type jobs aboard ships. During calendar years 1969-1970, the period with which this report is

concerned, blacks were sufficiently represented among the graduates of 24 Class "A" schools for inclusion in a bi-racial validity study of the BTB.

Method

The problem of test bias is complicated by the number of ways in which a test may be discriminatory. This study will concentrate on two commonly accepted definitions of bias, or lack of bias. The first is that of Cleary (1968) who stated, "A test is biased for members of a subgroup of the population if, in the prediction of a criterion for which the test was designed, consistent nonzero errors of prediction are made for members of the subgroup." Statistically, this type of bias is investigated by testing the slopes and intercepts of the regression lines for the majority and minority populations to determine whether they differ significantly. The method used for performing these tests was the one developed by Gulliksen & Wilks (1950). The second definition of discrimination investigated is that of the Department of Labor for the Equal Employment Opportunity Program and involves test fairness. In a section of Title 41 (1971) the following directions for assessing the validity of a selection test are given: "The relationship between the test and at least one relevant criterion must be statistically significant. This ordinarily means that the relationship should be sufficiently high as to have a probability of no more than 1 to 20 to have occurred by chance...A test which is differentially valid may be used in groups for which it is valid but not for those in which it is not valid." To determine whether the recruit classification tests comply with this standard, the BTB selection composites were validated against final grades in Navy schools separately for black and white samples.

Data routinely gathered for graduates and disenrollees from Class "A" schools formed the basis of the sample. BTB scores and racial information were obtained for students completing their training in 1969 and 1970. The data were sorted by race and school code to determine which schools had sufficiently large samples of blacks for a bi-racial analysis of possible selection test bias. Twenty-four schools (out of approximately 140) were selected because they had at least 19 black students among their graduates or academic disenrollees.

The total number (all "A" Schools) of white students with complete predictor and criterion variables was 104,683, while blacks numbered 2067. The records of blacks have not been isolated in previous BTB studies because of their small representation in the school samples and because the problem of possible test bias was not a salient issue. Now, however, Title 41 has shifted the burden of proof of nondiscrimination to the employer and the military services must determine whether or not their selection tests are biased.

The other variables used in the statistical analyses were:

Basic Test Battery (BTB). Scores on the BTB are reported as Navy Standard Scores having a mean of about 50 and a standard deviation of about 10 for an unrestricted recruit population. The tests used in this research were the General Classification Test (GCT), Arithmetic Reasoning Test (ARI), Mechanical Test (MECH), Clerical Test (CLER), Shop Practices Test (SP), and Electronics Technician Selection Test (ETST).

Armed Forces Qualification Test (AFQT). Scores on the AFQT are reported as percentiles with a minimum score of 10 established to indicate mental fitness for military training.

Final School Grade (FSG). The grade given upon graduation or disenrollment is most commonly a weighted sum of grades earned on daily and/or weekly quizzes, measures of practical proficiency, and the score on the final examination. It ranges from about 35 to 99 in its raw form and was standardized to a mean of 50 and a standard deviation of 10 for some of the analyses in this study.

Means, standard deviations, and correlations among the test variables and standardized FSG were computed for the two racial samples. The significance of the differences between statistics for black and white groups was determined. The regression lines for each BTB test and AFQT were plotted separately by race and tested for differences of errors of estimate, slope, and intercept using the method of Gulliksen & Wilks (1950). In addition, regression weights for all six BTB tests (using FSG as a criterion) were developed for the white students within each school. These weights were applied to the BTB scores of their black classmates to obtain the predicted FSG of each minority member. The difference between the actual mean FSG and the predicted mean FSG of the schools' minority students was determined and tested for significance.

In practice, aptitude for a Navy school is not determined by a score on a single BTB test or on scores weighted through regression techniques. Instead, a summed combination of two or three BTB tests, depending on the particular school, is used in the classification decision. Thus, the most relevant statistic for judging the effectiveness of the battery in school selection is the correlation between this composite and FSG. These correlations were computed separately for each race and tested for significance in line with the requirements of Title 41. The differences between the validities for blacks and whites within a rating were also tested.

Results

All of the racial mean scores on the BTB tests, the AFQT, and FSG differed significantly with the whites consistently performing

higher on the tests and in school. With one exception, that of CLER, the tests were significantly more valid for the whites also, even though the standard deviations were very similar. Although these results clearly show that the BTB and AFQT are better predictors of the school grades of white enlisted men than of black, the question of bias is still not answered.

The regression lines for MECH and CLER showed consistent positive bias; that is the school grades of blacks would be predicted to be somewhat higher when based on a majority sample than when based on a minority sample. The remaining regression lines presented a situation in which the grades of blacks scoring low on the tests are underpredicted, while the school performance of higher scoring blacks is overpredicted by white regression equations. For most tests, the two regression lines crossed below the mean test score of the minority sample. Thus, overprediction is more commonly the case than underprediction. On the AFQT, however, the lines crossed just above the mean test score of the blacks (57th percentile) so that over- and underprediction occur with almost equal frequency. Chi-square tests of the significance of the differences between the population variances, regression slopes, and intercepts were made for all tests. The results obtained for all seven aptitude tests demonstrated that the two racial populations were not homogeneous. The regression lines in each case are significantly different for whites and blacks. Thus, the assumption that these tests are related to the criterion variable in the same manner for blacks as for whites has not been supported.

Another index of bias is significant differences between school grades actually earned by the minority population and grades predicted by applying majority regression weights to minority test scores. If the predicted grades are lower than the actual grades, then capable minority members at the lower end of the test score distribution would be rejected and negative bias is said to exist. Twenty-four Class "A" schools, representing 19 different ratings, were involved in the multiple-regression analysis. The actual mean FSG of blacks was higher than the mean predicted FSG in 14 of these ratings and significantly so in five. In the five ratings in which black grades were lower than their BTB scores would indicate, the difference was not significant. Thus, if school selection were based on prediction equations developed from the majority student population (using all BTB tests), negative bias would be operating.

The most meaningful type of bias analysis is one which studies the tests as they are actually used in selection. For the Navy this means looking at the validities of the test combinations, as predictors of performance in the relevant schools, for black and white samples separately. Only 22 schools (18 ratings) were used in this analysis because the remaining two schools have varying selectors. When comparing the correlations between school selectors and school grades for black and white students linear-summed validities rather than

multiple correlations are used because test scores are simply added together to determine school eligibility (with the exception of the ARI+2ETST selector in which a weight of two is applied to one test). It was found that the operational selector composites were predictive of the school performance of white students at the .01 level of significance in every rating in the analysis. These same selectors failed to predict the grades of black students (above chance levels) in nine of the 18 ratings. From this analysis it appears that use of these particular test combinations violates Title 41 for the minority group. However, other BTB test composites could be used for school selection if it can be demonstrated that they predict the final school grades of black students. Therefore, the validities of all combinations of two BTB tests were determined. In all nine ratings in which the operational selectors failed to yield significant validities, prediction could be improved by using other combinations, significantly so for six ratings.

As reported earlier, the mean test scores of the blacks were significantly lower than those of the whites. There is a possibility that this factor, in addition to race, may have accounted for much of the differences in the validities of the test composites. To test this assumption, two samples of white students whose selection test scores matched those of the blacks were drawn from each school sample. The mean criterion scores were determined for each of these sub-samples. The overall means of the school grades were remarkably similar for the three groups, indicating that, in general, blacks and whites matched on aptitude level perform comparably in school.

Conclusions

There can be little doubt that the black and white samples differed significantly in their performance on the BTB and AFQT. All racial means were significantly different, six of the seven test validities were significantly different, and the hypothesis that the two samples were drawn from a homogeneous population was rejected. In the strict statistical sense adopted by Cleary, it has been demonstrated that bias exists in these tests. However, no consistent tendency was found for the tests to either overpredict or underpredict the school performance of blacks. Instead, the comparison of school grades of blacks and whites matched man to man on relevant test variables resulted in each group exceeding the other with equal frequency.

On the practical and legal question of the validity of the school selection composites, it was shown that for half of the ratings the selectors failed to predict the performance of black students at the .05 level of significance. However, the factors underlying this finding were not clear.

The ambiguity of the results precludes firm conclusions. The possible existence of selection bias, which has not been ruled out by

the findings, makes further research imperative. Under instructions from the Chief of Naval Operations, classification officers are assigning many more blacks than before to formal school training. If research on larger samples confirms the apparent differences between validities for blacks and whites, new selection criteria will have to be developed for minority recruits.

References

- Cleary, T. A. Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement*, 1968, 5(2), 115-124.
- Department of Labor. Proposed employment opportunity program. *Federal Register*, 1971, 36, No. 77.
- Guinn, N., Tuper, C. E., & Alley, W. E. *Cultural subgroup differences in relationships between Air Force aptitude composites and training criteria*. AFHRL-TR-70-35. Lackland AFB, Texas: Personnel Research Division, Air Force Human Resources Laboratory, September 1970.
- Gulliksen, H., & Wilks, S. S. Regression tests for several samples. *Psychometrika*, 1950, 15(2), 91-114.
- O'Leary, B. S., Farr, J. L., & Bartlett, C. J. *Ethnic group membership as a moderator of job performance*. Technical Report No. 1., April 1970, American Institutes for Research, Contract No. N00014-68-C-0341, Office of Naval Research, Department of the Navy.

SESSION II

Training

Human Factors

Training

Drugs and Rehabilitation

Training

CHAIRMAN: Major Dirk C. Prather

SERVICE TEST/EVALUATION OF MULTIMEDIA CDC 60300,
VEHICLE OPERATOR/DISPATCHER

Russell J. Hibler

Air Training Command

Evaluates the effectiveness of a multimedia CDC presentation compared to the same information presented in the conventional printed CDC. The conventional CDC 60330, Vehicle Operator/Dispatcher, was transcribed into a sound and slide presentation for use on the Raytheon 600 Mediamaster Trainer. Trainees using the multimedia presentation learned more, completed in less time, required less assistance and rated the method of presentation more favorably than the conventional CDC.

One of the most evident areas of the learning handicaps of low mental ability airmen was in the completion of home study materials, entitled Career Development Courses (CDCs) (ATC 1969a; ATC 1969b). These self-study courses are an integral part of the dual-channel upgrade program, where airmen gain knowledges and proficiencies essential to satisfactory job performance in their Air Force Specialty Codes (AFSCs). CDCs are written by specialists of the Air Training Command to accompany on-the-job training, and they are distributed by the Extension Course Institute of the Air University. As a group, low-ability airmen had difficulty completing the CDC materials. This was true, whether the CDC was designed to assist in the initial learning of a skill or the upgrading of a skill. Several steps have already been taken to more effectively meet these learner needs. Among the Air Force advances have been the testing of procedures to more efficiently adapt current CDCs to learner reading levels (Huff, 1970) and the trial of simplified written materials with more illustrations and audio supplements (Sellman, 1970).

The purpose of this study was to determine the instructional effectiveness of a multimedia CDC versus a conventional CDC, emphasizing the effects of the presentations on personnel of low mental abilities.

Method

Subjects

Sixty male airmen students on a preentry status into several airman basic courses at Sheppard AFB were selected. The airmen chosen were entering fields that either used vehicle operations or had similar AFSC prerequisites for AFSC 60300: Airman Qualification Examination (AQE) minimum score of 40 Mechanical. These airmen were divided into three ability groups of 20 men each based upon their Armed Forces Qualification Test (AFQT) scores. The selected airmen

had AFQT scores toward the centers of the AFQT Categories II, III, and IV and were defined as: high ability (AFQT percentile 70-87); medium ability (AFQT percentile 40-56); and low ability (AFQT percentile 12-26).

Treatments

Two variations of Volume I of the three volume CDC 60330, Vehicle Operator/Dispatcher, were used in this study.

Conventional. The standard CDC and a revised Volume Review Exercise (VRE) workbook utilizing 102 multiple-choice questions and paragraph references for finding the correct answers were used. Two AFSC subject matter experts answered any questions the trainees had while these materials were being used. Trainees were directed to read a chapter of the CDC volume, answer all of the review questions and check their answers with the referenced paragraphs. When they were unable to ascertain the correct answer, they asked one of the two subject matter experts for assistance. This process of reading, reviewing and asking questions when necessary was repeated for each chapter and is similar to standard CDC and workbook, or VRE, procedures. These materials were self-paced.

Multimedia. The subject matter used by the conventional CDC group was rewritten for audio presentation, recorded on audio tape and accompanied by 35mm color slides. The same review questions used by the conventional group were used for adjunct programming: as each lesson objective was presented, it was immediately followed by its multiple-choice review question, a pause for trainees to respond and the correct answer. This instructional procedure was repeated for all of the review questions at each chapter conclusion. This treatment was presented on the Raytheon 600 Mediamaster Trainer. This completely automated instruction was controlled by inaudible pauses that were programmed on the same magnetic tape as the audio portion of the presentation. Each student had a four-choice selector which recorded his responses to the review questions. This system can monitor a group of up to 60 trainees and requires a qualified operator. Dean (1966) offers a detailed description of this equipment and its air Force use. Two subject matter personnel were also present to answer questions. The multimedia trainees were group paced, making the time required for completion the same for all, or a total of 360 minutes.

The texts of the two treatments were compared, and the conventional CDC was found to be longer, to have fewer illustrations, but to be at the same level of reading difficulty as the multimedia version.

Criterion Measures

Criterion measures included difference scores (post-minus pretest) from a 75-item subject knowledge test, the number of minutes required

for completion, the amount of assistance required from the subject matter personnel and an opinion survey.

Procedure

The 20 trainees in each of the three ability groups were randomly assigned to the conventional and multimedia treatments, forming a 2X3 factorial design. This distribution of high, medium- and low-ability trainees enabled analyses of variance for both the difference scores and the completion times. The CDC treatments were administered in classrooms where the conventional CDC trainees worked at independent rates in a study hall style environment and the multimedia trainees were paced as a group by the Raytheon Trainer. Trainees proceeded through the materials during two-hour sessions which were held every morning until all men had reached completion.

Results

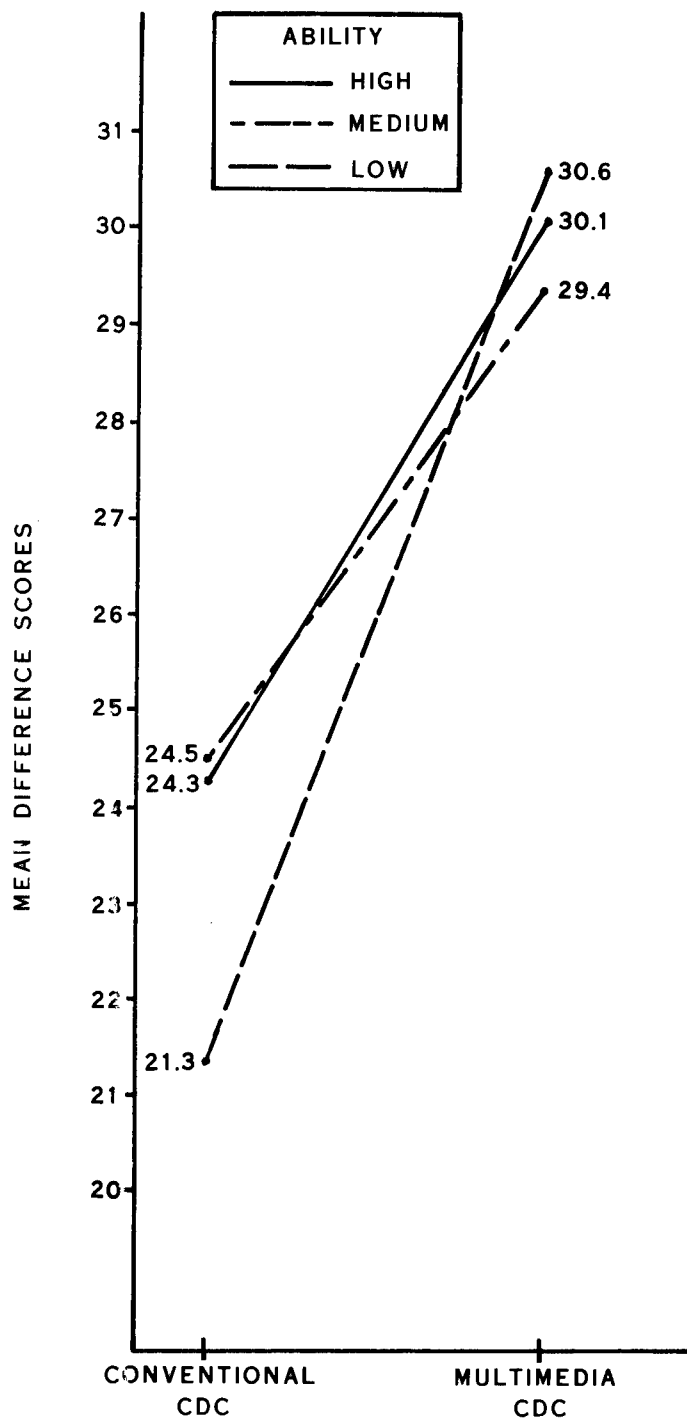
Knowledge Gains

The summary of the analysis of variance for the knowledge gained, or difference scores, appears as Table 1 and indicates the multimedia instruction yielding significantly higher gains ($p < .01$) in knowledge, see Figure 1.

TABLE 1
Analysis of Variance for Difference Scores

Source of Variation	Sum of Squares	<i>df</i>	Mean Square	F
Ability Group	17.500	2	8.750	.51
CDC Treatment	666.667	1	666.667	38.53**
Ability Group X CDC Treatment	54.033	2	27.017	1.56
Error	934.400	54	17.304	
Total	1672.600	59		

** $p < .01$



.14

Fig. 1. Mean difference scores for each CDC treatment and ability group.

Completion Times

The analysis of variance for completion times, Table 2, shows the conventional CDC trainees took significantly longer ($p < .01$) to completion than the multimedia trainees, see Figure 2.

TABLE 2
Analysis of Variance for Completion Times

Source of Variation	Sum of Squares	df	Mean Square	F
Ability Group	15085.833	2	7542.915	.80
CDC Treatment	72356.917	1	72356.917	7.56**
Ability Group X CDC Treatment	13799.323	2	6899.667	.73
Error	510562.500	54	9458.861	
Total	611804.573	59		

** $p < .01$

Assistance Required

The numbers of questions asked the subject matter experts were as follows: conventional CDC, 141 questions asked; multimedia CDC, no questions asked. During the conventional CDC instruction, the subject matter experts assisted for the full time of the study sessions and were properly utilized.

Trainee Attitudes

On each of the six items of the opinion survey, as well as the total of the items, treatment group differences were not significant. The written comments of both groups were laudatory and emphasized the attributes of each presentation method. These comments indicated that the multimedia CDC trainees had more to say and were more favorable in their opinions than the conventional trainees.

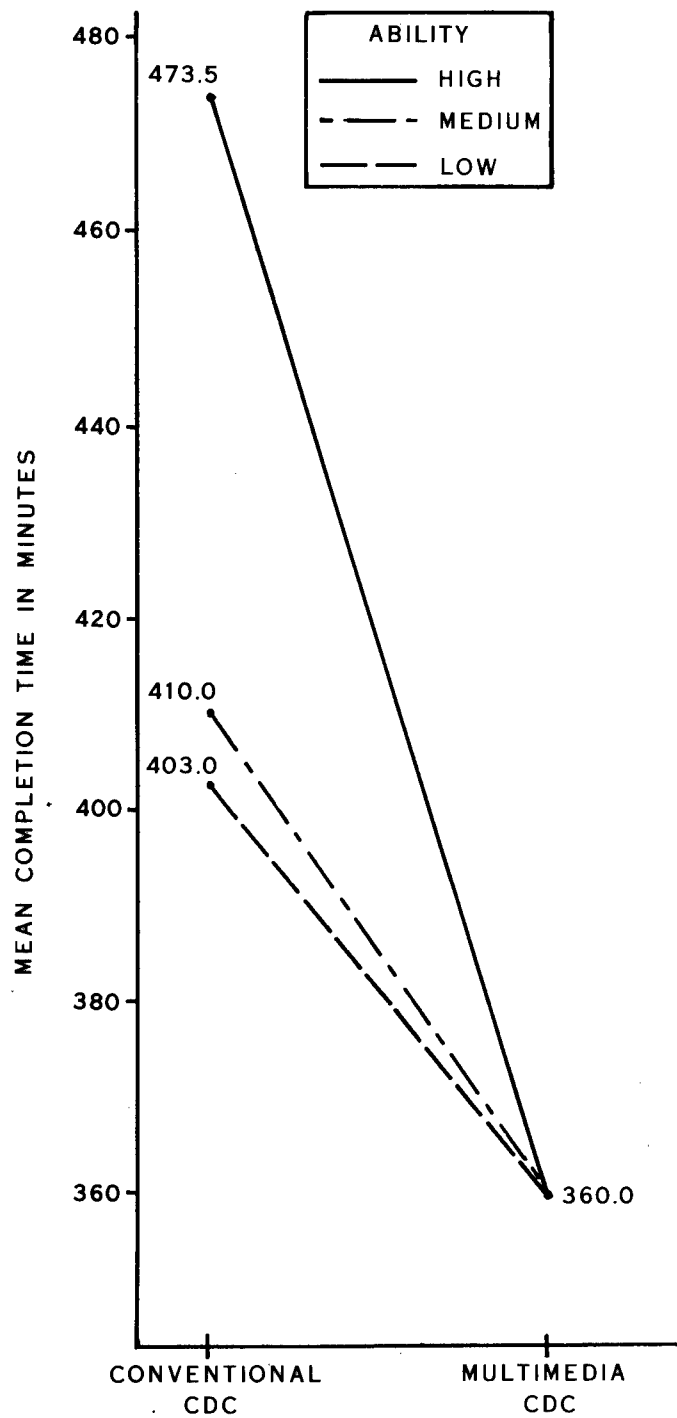


Fig. 2. Mean completion times in minutes for each CDC treatment and ability group.

Discussion

Knowledge Gains

Trainees of low ability, and of medium and high ability as well, learned more from the audio and 35mm slide, or multimedia, presentation than was learned from the conventional CDC presentation. Gage (1967) and Cronbach (1967) attributed differences between treatment effectiveness to the adaptation of treatment variables to students' individual differences. In this study, circumventing a strong dependence on reading and, particularly on studying, through the oral instruction and review, contributed to this overall effect. It is assumed that the greater number of illustrations and shorter script of the multimedia CDC made it more explicit and contributed to its trainees' superior performance. Within the conventional CDC instruction, the differences among ability group performances were not statistically significant, but were in the direction of the low-ability trainees gaining less from the conventional CDC instruction than the higher ability trainees. This finding was in the direction of other studies (ATC 1969a; ATC 1969b). On the multimedia task, the low-ability trainees performed at least as well as the other trainees when they used the multimedia instruction. It is through this type of gain in instructional effectiveness from one treatment to another for a specific group that students can be prescribed the most effective training methods for their individual learning abilities (Bracht, 1971).

Completion Times

The significantly shorter ($p < .01$) amount of time required by the group-paced multimedia trainees is consistent with findings of Federico (1971). He found that, in a resident training course and with Air Force students, multimedia instruction took less time to completion than self-paced (programmed) instruction. The direction of higher ability students taking longer, although not statistically significant, supported Klausmeier & Goodwin (1966) in that high-ability students value and have developed study habits for written or the conventional CDC mode of presentation and, therefore, retained the material longer than low-ability groups.

Trainee Attitudes

The overall similarities in acceptance for both instructional modes, in part, can be explained by a novelty effect, since this was the airmen's first exposure to Air Force technical training. The less structured, written comments did indicate differences between treatment acceptances and may be a more appropriate type of measurement. Federico (1971) also found themes or comments to show that Air Force students had higher opinions of audiovisual instruction than self-paced, written (programmed) instruction. In the present study,

students indicated that their acceptance of multimedia instruction was based on its characteristics themselves, color slides, adjunct programming, etc.

Conclusions

The multimedia (sound and slide) presentation of information in the Career Development Course 60300 is more effective than the conventional printed media. Compared to students who used the printed media, students receiving the multimedia presentations learned more, took less time, required less help and reacted more favorably to the training.

References

- ATC Project Report 69-1, *Survey*, Project 100,000 DDA/CDC. Sheppard Technical Training Center, Sheppard AFB, Texas, May 1969. (a)
- ATC Project Report 69-2, *Test Graduate Evaluation Report*, Project 100,000. Sheppard Technical Training Center, October 1969. (b)
- Bracht, G. H. Experimental factors related to aptitude-treatment interactions. *Review of Educational Research*, 1970, 40(5), 627-646.
- Cronbach, L. J. The two disciplines of scientific psychology. *American Psychologist*, 1957, 12, 671-684
- Dean, P. Systematizing traffic safety instruction. *USAF Instructor's Journal*, Vol. IV, No. 2, Fall 1966, 30-35.
- Federico, P. A. *Evaluating an experimental audiovisual module programmed to teach a basic anatomical and physiological system*. AFHRL-TR-71-37, Brooks AFB, Texas, 1971.
- Gage, N. L., & Unruh, W. R. Theoretical formulations for research on teaching. *Review of Educational Research*, 1967, 37, 358-370.
- Huff, K. H., & Smith, E. A. *Reliability, baseline data and instructions for the Automated Readability Index*. AFHRL-TR-70-14, Brooks AFB, Texas, 1970.
- Klausmeier, H. J., & Goodwin, W. G. *Learning and human abilities*. New York: Harper & Row, 1966.
- Sellman, W. S. *Effectiveness of experimental training materials for low-ability airmen*. AFHRL-TR-70-16, Lowry AFB, Colorado, 1970.

DEVELOPMENT OF AN ADVANCED TRAINING RESEARCH SIMULATION SYSTEM

Don R. Gum, Patricia A. Knoop, James D. Basinger,

Israel M. Guterman, and William L. Foley

Air Force Human Resources Laboratory

An advanced training research simulation system is being developed which will be located at an active Undergraduate Pilot Training (UPT) Base (Williams Air Force Base) and utilized with UPT students in an experimental program of pilot training research to answer some of the many questions concerning training in simulators.

The development effort is being pursued under Project 1192, Advanced Simulation in Undergraduate Pilot Training (ASUPT). The formally stated objectives of the total ASUPT Program, including system development and utilization are (a) to enhance pilot training within the Air Force through the application of recent technological advances in simulation (b) to demonstrate the maximum effective utilization of simulators in Air Force Undergraduate Pilot Training, and (c) to define the future generation UPT ground training equipment and simulators.

The role of simulators in UPT is an area essentially untouched from a training research point of view. Basic principles of human learning and skill acquisition suggest that simulators could find their most effective application in the training of pilot candidates.

In order to investigate the limits of the latest simulation technology and to define methods and techniques for the maximum utilization of this technology in an on-going UPT program, an appropriate simulation system incorporating the latest technology built around the present training aircraft must be available.

Since the present ground trainers utilized in UPT are only instrument and procedures trainers with low flight fidelity and no motion or visual simulation, an appropriate advanced training research simulation system must first be developed. That is the purpose of the development effort described in this paper. A companion paper, *Application of the Advanced Simulation in Undergraduate Pilot Training (ASUPT) Research Facility to Pilot Training Programs*, describes how the simulation system is to be utilized.

The total ASUPT Simulation System is comprised of three major components: (a) two basic T-37B simulators, (b) two wide-angle infinity visual displays, and (c) a shared visual computer image generator.

Basic Simulators

The basic simulators are modeled and programmed to simulate ground operations, normal flight conditions, emergency flight conditions, aerobatic flight, formation flight, and post stall and spin in a high fidelity manner. The cockpits include faithful reproductions of in-cockpit sights, sounds, and control feel to the maximum extent allowable by the state-of-the-art and simulation realism versus functionality compromises.

Mathematical Modeling

The mathematical modeling is complete and high fidelity in most areas and implemented in such a manner that it can be systematically degraded by a researcher for studies of transfer of training as a function of fidelity.

The aerodynamic models for both the simulator and the test criteria development are based on a rigorous set of aerodynamic equations. Also to verify the validity of the aerodynamic data, the test criteria model performance is being verified against Air Force flight test data. Coefficient data have been developed for angles of attack ranging through $\pm 90^\circ$ and sideslip angles ranging through $\pm 180^\circ$. This coefficient data is applied to the standard rigorous aerodynamic equations as the means for simulating stall, post stall, and spin in the simulators.

The model is implemented in such a way that researcher degradation can be controlled by (a) specifying trigonometric function accuracy either the complete function, small angle approximation, or deletion and (b) altering multiplier coefficients to modify the effect of terms of the aerodynamic equations and the effect of individual aerodynamic coefficients. The simulation can also be degraded through the off-line modification of the aerodynamic coefficient function generation. A function generation compiler program which will allow the researcher to specify the accuracy to which a given function is to be represented will be provided to facilitate this method of degradation. Also, in order to realistically simulate the conditions experienced in formation flight, modeling of jet wake and downwash from the lead aircraft is included.

Motion and Force Simulation

The motion and force simulation is accomplished through a combination of a six degree-of-freedom synergistic motion system and a sustained "g" seat.

The motion system provides the on-set acceleration cues along and about the three aircraft axes. The system being synergistic has the platform supported by six active hydraulic actuators. Also included

are six passive safety actuators providing complete mechanical redundancy in case of system failure. The system is essentially a hydraulic position servo driven by commanded leg or actuator lengths computed by the motion system mathematical model. The system is designed for a 23,000 pound load carrying capacity. The motion system performance capabilities are shown in Table 1.

The sustained accelerations are simulated through a combination of an activated lap belt and compartmentized air inflatable seat, back, and thigh cushions. The seat cushion is composed of 16 individual air activated compartments; the seat back is composed of 9 individual air activated compartments; and each thigh cushion contains three individual air activated compartments. The lap belt is activated by two air controlled pistons, one on each side.

TABLE 1
Motion System Characteristics (Nonsimultaneous)

	Position	Velocity	Acceleration
Heave	+39", -30"	$\pm 24''/\text{sec}$	$\pm 1g$
Lateral	$\pm 48''$	$\pm 24''/\text{sec}$	$\pm .6g$
Longitudinal	$\pm 48''$	$\pm 24''/\text{sec}$	$\pm .6g$
Pitch	+30°, -20°	$\pm 15^\circ/\text{sec}$	$\pm 114^\circ/\text{sec}^2$
Roll	$\pm 22^\circ$	$\pm 15^\circ/\text{sec}$	$\pm 114^\circ/\text{sec}^2$
Yaw	$\pm 32^\circ$	$\pm 15^\circ/\text{sec}$	$\pm 114^\circ/\text{sec}^2$

The sustained accelerations are imparted to the pilot through the cushions by varying the orientation and contour of the seat and back planes. The variation of the seat and back planes alters the direction of the force vector and the variation of the contour alters the contact area thereby altering the pressure applied to certain parts of the body creating the illusion of a change in the magnitude of the force vector.

Computation

The computer system being used to drive both simulators is composed of a single SYSTEMS 86 central processor unit (CPU) with 98,304 words a core memory. Included also are the usual peripheral devices such as: a teletypewriter, a line printer, a card reader, a disc, two magnetic tape units, and a digital plotter. The computer system is interfaced through an analog and discrete linkage system with the simulator cockpit and instructor/operator station (IOS) controls and instruments, and directly with the four digitally controlled CRT displays at the advanced IOS.

The ASUPT software and its execution are centered about the SYSTEMS 86 Real Time Monitor (RTM). This is a disc-oriented multi-programming monitor system providing 64 software priority levels for control of both foreground and background tasks. Standard features of RTM include interrupt and trap processing, reentrant monitor services available to both foreground and background jobs, file management, batch processing, and program overlays.

The simulation load is organized as a single foreground resident task under RTM with its own executive for processing the jump list. The simulation resident image in core consists of a task service area, the executive, and the simulation resident image in core consists of a task service area, the executive, and the simulation programs, including advanced instructional provisions, CRT handles, and real time I/O handlers. The majority of the simulation programs are written in FORTRAN. Beyond the RTM and the simulation load, core is partitioned into a data pool area and background space.

Unique to ASUPT, as a full-fidelity simulation system, is the provision for foreground/background operation. This is made possible by the utilization of a sophisticated monitor (SEL RTM) which permits background job execution, including compilations and assemblies, to be interrupted by the real time load (foreground task) and resumed during the spare time of each real time frame. This feature provides considerable flexibility by enabling programs unique to various simulator experiments to be prepared and executed with the simulation program. In addition, it will facilitate simulator modification by allowing changes to be compiled and debugged in the background with no simulator downtime.

Instructor/Operator Stations

Six stations are provided for the two simulators: one conventional station, one combined advanced/conventional station, two in-cockpit instructor stations, and two in-cockpit student stations. The combined station consists of a conventional station mated with an advanced station and may be used in several configurations. The two cockpits may be controlled from their respective in-cockpit instructor stations, the advanced portion of the combined station, or a

conventional station, as selected by a master mode control on the advanced station. One of the cockpits may additionally be controlled from the combined station.

The conventional station is a standard instructor/operator station using repeater instruments and traditional I/O common to instrument trainers of the past and most mission simulators of the present.

The right hand portion of the combined advanced/conventional station is a conventional station with the addition of some CRT controls. This portion can be used as a conventional station or as a combined advanced/conventional station. The latter mode augments the conventional station with one of the CRT displays on the advanced portion. This can be used for GCA, cross country and aerobatic maneuver plotting and monitoring.

The left hand portion is an advanced station comprised essentially of 4 CRT displays (2 alphanumeric and 2 graphic), push-button switches for CRT assignment and contents-control, and a keyboard. Any of the 4 CRTs can be assigned on demand to either cockpit. A number of alternative CRT pages are provided for call-up on any CRT compatible with the type of page (alphanumeric or graphic).

Alphanumeric pages, collectively, replicate the data provided at the conventional station and provide user interface with the advanced instructional provisions. Hard copy of any alphanumeric page can be obtained on demand. Graphic pages include cross country, GCA, formation flying, and a spatial display. The latter provides a 3-dimensional view of maneuvers, and the apparent viewpoint can be altered by rotating the image about any axis via panel controls. A control stick is also located at the advanced station for use in conjunction with the formation flying display in flying the lead aircraft.

An in-cockpit instructor station is located in each cockpit to the right of the instructor's seat. These stations consist essentially of a CRT display, keyboard, and control switches. Any alphanumeric CRT page available at the advanced station may be called up on the in-cockpit station. This gives the in-cockpit instructor access to all advanced instructional provisions, including record/playback capabilities.

A student station is located in each cockpit to the left of the student's seat. These stations consist essentially of pushbutton and thumbwheel switches and are used primarily for student-directed training. The in-cockpit CRT is viewable by the student when the right seat is empty. Using this CRT and his station controls, the student may select exercises or maneuvers for practice, and request automated demonstrations.

Advanced Instructional Provisions

Advanced Instructional Provisions (AIP) are included in ASUPT to (a) make conduct of training independent of variance in instructional technique insofar as possible, and hence provide the standardization essential in conducting research, and (b) provide a basis for evaluating the impact upon training of automated instruction, student-directed training, and (when feasible) adaptive training.

Seven AIP are provided, each of which is a correlate to one or more functions that a simulator instructor, operator, or experimenter (hereafter referred to collectively as IOE) performs in conducting training. The IOE normally begins training by briefing the student and demonstrating tasks to be practiced. One AIP provided, therefore, is for *Automated Demonstrations*, enabling exemplary performances to be prerecorded, with aural comments, and used during training in a fast, real, or slow time mode.

As students' skill varies, the IOE changes task difficulty to present a challenge to the student commensurate with his abilities. The IOE also inserts malfunctions to provide training in emergency procedures. Both *Variation of Task Difficulty* and *Automatic Malfunction Insertion* are provided as AIP, the former including variation of system dynamics, motion cues, and environmental parameters.

As instruction proceeds, the IOE monitors the student's performance. In ASUPT, this is enhanced by providing *Instructor Feedback* consisting of a repertoire of CRT pages for on-demand call-up and a capability for plotting on the CRT actual versus criterion performance. Criteria may be prerecorded or computed from a recorded demonstration.

The IOE records observations about performance for record-keeping and student debriefing and provides feedback to the student during training. In ASUPT, automatic *Data Recording* of simulation parameters is provided, with recording rates and output devices selectable. Automated *Student Feedback*, using either aural or visual media, is also provided. Aural feedback is accomplished using a computer controlled "speech-maker" unit. Visual feedback is supplied on the in-cockpit CRT or inset in the simulated visual scene.

Finally, the IOE makes decisions about the direction of training and sequences tasks to be practiced accordingly. *Automated Task Sequencing* is provided, enabling IOE or student manual selection of the next tasks, or an automated selection sequence that has been pre-specified or is computed based on task difficulty and importance.

For effective utilization, many AIP are dependent upon some method of assessing student performance. The IOE performs this function using a subjective rating method. Techniques that are more objective and reliable are being developed under a separate effort to be added as an AIP at a later date.

The AIP are implemented as resident modules in the simulator load. Some permit direct control and can be cued into execution from the advanced instructor/operator station. For automatic execution of AIP, a "pre-programming" capability is provided, which is a technique for creating a computer-managed training program by specifying parameters for each AIP and organizing them into a sequence of execution. Pre-programming Statements may be prepared on cards or inserted through a CRT keyboard. In case of the latter, the CRT is employed to guide the user through the steps he must take to create, modify, or delete pre-programmed exercises, which are maintained in a temporary disc file.

Visual Systems

The ASUPT visual simulation characteristics, such as field-of-view, resolution, area of coverage, and altitude and attitude range requirements, are based on the UPT aircraft to be simulated and a detailed analysis of each major UPT flying task. The most difficult characteristic to achieve for ASUPT is the field-of-view which is approximately $\pm 120^\circ$ horizontal by $\pm 120^\circ$, -40° vertical. Of all the various visual simulation techniques investigated, only the mosaiced in-line infinity display driven by a multi-channel Computer Image Generation (CIG) system has the capability for fulfilling the majority of the visual simulation requirements for the varied UPT flying tasks. The visual system will provide the extracockpit environment for taxiing, take-off, approach and landing, airwork and aerobatics, and formation flying.

Visual Displays

To achieve the wide field-of-view requirement, the total display system is formed from seven pentagon-shaped display channels. These channels are mosaiced together to form a partial dodecahedron shell surrounding the cockpit. Each display channel is a separate in-line infinity display. An infinity display is a type of image relaying system in which the image appears to originate at infinity or at a far distance from the viewing point. The display is characterized as an in-line display due to the configuration which permits the collimating reflective optics to have an optical axis coincident with that of the input cathode ray tube (CRT). It is this in-line configuration that permits the display channels to be mosaiced together to form a continuous wide field-of-view. However, there is a price to be paid for this in-line feature, which is the loss of display optical efficiency. The display optics are approximately 1% efficient which necessitates the use of high-brightness CRTs. This, in turn, eliminates the possibility of providing a color display since color CRTs of the required brightness are beyond the state-of-the-art. The displays will therefore be monochrome using CRTs with P-20 phosphor.

The spherical beamsplitter, the primary element of the display optics, has a radius of curvature and distance from the viewing point of 48 inches. The CRTs are 36 inches in diameter with a 24-inch radius faceplate. These are the largest CRTs ever to be developed. The display brightness at the pilot viewing point is to be 6 foot-lamberts and the display is to have a resolution of 1000 TV lines.

Computer Image Generator

The CIG system generates a video signal which, when displayed by the visual display described above, presents a simulated visual scene to the student pilot. The CIG video signal, generated by a special digital computer, is a signal similar to that generated by a television camera. However, since a camera is not employed, none of the television camera type constraints exists. The CIG generates the video in real time for visual scenes having the following primary capability (which are normally camera constrained):

1. Exact perspective, since it is computed
2. Moving objects, such as a lead aircraft for formation flying
3. Quick visual environment change or modification
4. Unlimited attitude position and rates
5. Large area of flight coverage
6. No generation registration problems for multiple channels.

The CIG image, as the pilot views it on the display, consists of surface patterns or objects formed by planes of different brightness levels bounded by straight lines or "edges." The number of edges in a scene is a relative measure of image content and CIG system performance. Since scenes in the real world are not constrained to representation by straight lines or edges, a CIG system with a finite edge generation capability tends to generate a somewhat stylized presentation. The degree of stylization is inversely proportional to the edge generation capability of the CIG system. The required image content for training purposes is one of the variables of the ASUPT experimental program.

The CIG system stores a simulated visual environment model on a magnetic disc. This model is in numerical form and can be quickly and changed, modified, or amended. The CIG system, using aircraft position data from the simulator, extracts the portion of the environment model which the pilot can see and stores this into working storage. As the pilot flies in the environment, the working storage is continuously being up-dated according to the current aircraft position. Thus, the CIG system only processes the portion of the environment model

which the pilot sees and the total environment model can be several orders-of-magnitude larger than the model which the CIG can process in real time.

The CIG system then processes the visual model in working storage according to the simulated aircraft linear and angular position, as supplied by the simulator computer. This processing transforms the three-dimensional visual model into a two-dimensional display plane model. In the ASUPT CIG system, there are seven display planes since there are seven display channels. This display plane model is further processed to provide the occulting of farther objects by near objects.

Up to this point, the CIG system has been working in an edge format: storing, retrieving and transforming. This edge format is next converted into a digital scan line format. In this format, the brightness level of each part of the scan line is in digital form. This scan line information is then converted into a video signal by a high-speed digital-to-analog converter. This signal is then distributed to the various displays for viewing by the student pilot.

The CIG generated scenes are formed by the following basic visual elements:

1. Surface plane
2. Objects
3. Moving object
4. Sky
5. Perspective lights
6. Point lights
7. Special purpose lights

The implementation of the above elements depends upon the environment model, type of mission, and time of day.

To improve image quality, two techniques, edge smoothing and continuous shading of surfaces, are employed. The edge smoothing feature provides a gradual transition across an edge between adjacent shades of gray approximating the imagery produced by a television camera. The continuous shading of surfaces capability permits the generation of imagery representing curved surfaces. This will be used primarily on the lead aircraft for formation flying.

The major performance characteristics of the ASUPT CIG system are presented below:

1. Number of edges to be displayed: 2000
2. Model area capability: 1250 x 1250 nautical miles
3. Environment model to be delivered will include:

- a. Williams AFB

- b. All T-37 contact practice areas
- c. Headpin auxiliary airport
- d. Formation lead aircraft
- e. A 50nm perimeter around a, b, and c above.

4. Number of edges to be delivered in the environmental model: approximately 100,000 edges.

5. Total environmental model storage: 600,000 edges.

6. Television standards: 1023 scan lines and 30 frames per second.

7. Expansion capabilities:

- a. Color television outputs
- b. Additional 2000 edges to be displayed

8. Operation: The CIG can supply independent visual scenes to two simulators simultaneously with any ratio of edges between the two for a total of 2000 edges.

System Development

The three major components of the total ASUPT Simulation System are being developed by different contractors: the Basic T-37B Simulators by the Simulation Products Division of the Singer Company; the Visual Displays by the Farrand Optical Company under a sub-contract to Singer; and the Computer Image Generator by the Space Division of the General Electric Company. The Singer Company also has the responsibility for integrating the total system. The two Basic Simulators and Visual Displays are to be delivered in June 1973, the Computer Image Generator is to be delivered in November 1973, and the total system is to be integrated and operational in January 1974.

This system when completed will represent, with respect to advanced training features, the largest and most sophisticated simulation system ever developed. It will be the first time that such a system has been developed as a behavioral science tool strictly for application to pilot training research.

APPLICATION OF ADVANCED SIMULATION IN UNDERGRADUATE
PILOT TRAINING (ASUPT) RESEARCH FACILITY
TO PILOT TRAINING PROBLEMS

James F. Smith

Air Force Human Resources Laboratory

Presents a brief discussion of the evolution of simulation and cites two periods wherein flight simulation studies or related events led to the procurement of ASUPT. Describes the research capabilities included in ASUPT. Notes problems which must be considered in devising an experimental design. Discusses research strategies which are being considered in developing the ASUPT research plan, to be ready by early 1974.

Wilbur and Orville Wright learned flying by trials and numerous errors. Lt. Benjamin Foulois, the first military aviator taught himself with the help of a "correspondence course" from the Wright brothers. From this start, today's pilot training methods and programs evolved. Training devices and simulators (of sorts) also became an early part of flying training technology. For example, the first Link Trainer was developed in 1929. However, it was not until World War II that such trainers were accepted and procured on a production basis (Kelly, 1970). These instrument trainers all had motion.

In the early 50s, due to changes in marketing policy by industry and Air Force naiveté in trainer requirements, motion disappeared and a new series of computer controlled fixed base devices arrived on the scene. These included C-11A, T-4, and T-7/26 instrument trainers.

However, flight simulator technology was gaining momentum, and even though the use of trainers had its start in World War II, the "golden age" of simulator usage probably dates back to 1950-1954 when personnel at a USAF pilot training research laboratory located at Goodfellow AFB¹ demonstrated that a 1-CA-2 trainer with motion and visual could be used to replace 30 aircraft hours out of a 130-hour syllabus (Flexman, 1954).

Even with these positive results, there was no great impetus by the Air Force to adopt simulation. Instead, a number of questions were raised. "Are three degrees of motion really needed? Does this

¹ These included Hagin, W. A., Flexman, R. E., Houston, R. L., Matheny, W. G., Smith, J. F., Brown, E. L., and Boyle, D. J., most of whom are still active in pilot training research.

simulator fly like the aircraft? Is it worthwhile to use all this space to provide a visual scene? Can the results be repeated in an operational situation? Wouldn't it be better to buy more aircraft and fly more hours?" And so on...

Shortly thereafter, USAF research mission priorities changed and the Goodfellow laboratory was abandoned. However, the research results reported were later used by one of the research personnel to assist in selling a total simulation package to American Airlines.

From 1954 to 1967 USAF research in the development and design of new simulation equipment continued but operational evaluation of the devices in pilot training programs was seriously curtailed. The demise of such applied research seems to stem from the fact that to obtain subjects and to relate to operational problems, the research activity must be conducted at a training installation. However, such a program is usually long term, costs money, and will, if innovative, surely interfere with the standard syllabus. As a result, if research results cannot be shown to be immediately applicable, the research program is the first to go; five such USAF research installations met this fate between 1953 and 1958.

Since 1967, four significant events have occurred which should probably be termed a renaissance period in the field of simulation. First, the airlines applied the latest simulation technology including visual systems to their training programs with substantial cost savings. Second, the Army has reported that use of the Synthetic Flight Training System (SFTS) resulted in a reduction: of helicopter instrument training hours from 60 to 6.5; in total training hours from 86 to 49; and in course length from 12 weeks to 8 weeks (Caro, 1972).

Third, the USAF initiated two exhaustive contractual studies to develop a long range plan for use in UPT during the 1975-1990 time frame. Following these, a USAF mission analysis team consolidated the results into a recommended USAF position. The results of this analysis, not yet released officially, are conservative but in general corroborate the study results obtained by Caro.

16

And fourth, in 1967 the USAF initiated development of a research facility which could provide answers to flight simulation equipment design and training application questions for use in future pilot training. This equipment will provide the latest design in simulation domains for both instrument and contact pilot training. The equipment is planned to be ready for research in January 1974 (Gum, 1972).

As must be apparent from Mr. Gum's description of ASUPT, this system will permit an in-depth examination of all research domains examined in earlier pilot training studies, as well as increased attention to the use of bisual scenes and the investigation of inter-actions among all these domains. In fact, ASUPT will provide the United

States Air Force with the most advanced and most complex and versatile pilot training research facility in the world today. It is these last two factors, complexity and versatility, which generate the critical planning effort which must be accomplished prior to initiating research and which is the subject of the rest of this paper.

ASUPT Research Capabilities

Mr. Gum described the physical and design characteristics of the five major research domains of ASUPT. What remains to be discussed are the kinds of research issues which may be addressed in each of these domains.

a. The *Advanced Instructor Station Displays* will permit examination of what kinds of information are most effective for instructor usage in his role as training manager.

b. The *Advanced Instructional Features* will permit adoption of proven learning theory in a student centered atmosphere to enhance student achievement.

c. The *Fidelity* of the simulator may be varied with respect to dynamic performance, environmental conditions, and cueing to examine effects on S learning.

d. The variable of 6 degrees of freedom (DOF) *motion system* (G-seat included) will allow investigation of the perennial questions concerning the requirement for motion in most categories of pilot training.

These questions, to be answered for different levels of training and different categories of maneuvers, include at least the following: When is motion required?; How many DOF?; What are the most effective response rates?; What is the required excursion for each motion parameter?; and What drive signal philosophy results in most effective cueing?

e. The computer image generated (CIG) visual system and wrap-around display will permit investigation of continual recurring issues with respect to visual requirements for different categories of maneuvers and levels of experience. These issues will include at least a determination of the most effective field of view (FOV), picture content, scene resolution and examination of the effects of distortion. There are other visual problem areas noted by Wolff (1972) which cannot be addressed due to equipment limitations.

There is one additional problem area which will be examined using ASUPT and that is the *interaction* between motion and visual cueing (Matheny, 1972). Specifically, what are the motion requirements if a visual scene is added?

While the questions and research issues enumerated above are by no means all inclusive, they do emphasize the complexity, capability, and versatility of the ASUPT facility. In addition, the research domains identified include research capability which encompasses most academic disciplines. For example, behavioral scientists may concern themselves with visual perception, kinesthetic and vestibular cueing, and perceptual motor skills research. Capabilities for presenting with changing instructional information and applying learning concepts in a student oriented learning environment, and reexamining the Ss instructor's role with respect to managerial functions and information requirements should be of interest to educational psychologists. Human factors engineers also have available a wide range of opportunities.

Factors Affecting Research Issue Priorities

While the complexities and capabilities of ASUPT present great opportunities, the same features also create problems. One of these, which will be discussed briefly, is planning for most efficient use of the system. Obviously, the research plan must address realistic Air Force information requirements as well as "pure" research objectives. To tread this fine line is difficult but necessary to remain in existence.

Some information is already available which will assist in prioritizing the simulation research. The Army's SFTS facility has similar capabilities in two of the five research domains noted earlier, i.e., instructor station design and advanced training features. Considerable information should become available over the next two years which could reduce significantly the questions remaining to be answered in these domains. In addition, numerous studies are available which indicate high fidelity increases transfer of training. Since maximum transfer of training is desired, this issue will likely lose significance by 1974.

The remaining research domains, motion and visual, will surely retain their high priority for several years; partially because there is so much to examine and so little equipment available, and partially because they represent significant dollar costs in device procurement. Of these two research domains the visual system would seem to be the more important because: it is the most costly element of simulation; it represents the greatest potential for flight time savings (approximately 70% of the UPT program is contact work); and is the one domain in which ASUPT is the only existing or planned device possessing this research capability.

Using the above comments to arrive at a prioritized research program, it is next necessary to select an experimental design. While firm decisions have not yet been made, considerable thought has been given to this problem, which I would like to highlight briefly.

Selection of an Experimental Design

There are two principle research strategies which would appear feasible for use with ASUPT. First, an empirical approach could be applied. In this approach all variables of each research domain could be applied. In this approach all variables of each research domain could be examined in one experimental design using an analysis of variance (ANOVA) statistical treatment to examine the impact of each variable and the resultant interactions.

Second, a pragmatic approach could be used in which some level of simulation is defined for all domains and a study conducted to determine the effectiveness of this device.

Empirical Approach

To construct a factorial design for use in this project, it is necessary to select the variables to be examined in each research domain and the number of levels or combinations to be considered. We have already determined the first three domains will be optimized based on other research. This leaves visual and motion. The variables to be examined in the visual system have been determined to be FOV, resolution, distortion, and content. We must now estimate levels or possible combinations for each of these.

For FOV, there are seven windows of which it is assumed the front window is always needed. The result is 64 possible combinations of visual scenes. For the variables of resolution and contact, let us assume two levels of operation, i.e., one at maximum and one at 50% of that capability. For distortion, let us assume two levels, i.e., no distortion and one JND level of distortion. A summary of the levels for each of these variables is as follows: FOV - 64; resolution - 2; content - 2; and distortion - 2.

The remaining research domain to be considered is motion. While ASUPT is limited in excursions to those provided to a 60" stroke system, within that range we have the capability of assigning given levels to selected problem areas. These include: DOF 1 - through 6; extensor excursion - long vs short; drive philosophy - clipped vs proportional; rough air - on or off; and G-seat - on or off. A summary of the levels for motion variables is as follows: DOF - 64; excursion - 2; drive philosophy - 2; and G-seat - 2.

From the above, the number of cells required for ANOVA is:

$$\text{ANOVA} = 2^{(7)} \times 64 \times 64 = 524,288$$

Using 5 \$ per cell for reliability, some 2.6 million \$ would be required; Williams AFB produces approximately 400 \$/year. Hence, with

this design, assuming UPT stayed the same, the student population remained comparable, and the researchers don't give up, we could provide answers to these questions by year 8525!

An alternative would be to retain the factorial design, but to further reduce the number of controlled variables by judicious use of subjective opinion. A second alternative would be to use response surface fitting techniques using selected ordered variables as the basis for selecting the first set of experimental conditions. Followup efforts could then be used to refine the preliminary finding or to pursue special areas of interest.

Pragmatic Approach

From our customer's view, the adoption of a pragmatic approach has more appeal. For example, all five domains of ASUPT could be subjectively maximized and this total system tested in the UPT program. A better strategy may be the combination of a pragmatic and statistical approach. This would involve estimating the region of levels of best response and then reestimating the optimum levels through response surface fitting. Even with this approach, it may be necessary to study simultaneously, a limited number of variables. Another variation would be to configure ASUPT to any specified simulator design and then run a short term study to estimate the contribution of this device configuration to pilot training. Obvious advantages of the pragmatic strategy are:

- a. Research findings from other studies can be adopted to permit early optimization of ASUPT capabilities.
- b. Variations in more costly domains of simulation such as visual FOV display requirements can be addressed early for cost effective application.
- c. The research program could be interrupted to address operational problems of urgent interest outside of UPT without loss of data.

Summary

In this paper, I reviewed briefly the potential of the ASUPT system and identified five significant research domains. These are: advanced instructional features, instructor station design, fidelity levels, motion system parameters, and visual system parameters. Of these five domains, visual and motion systems are anticipated to be of greater significance in 1974.

The advantages and disadvantages of two research strategies, empirical vs pragmatic, were discussed. It is concluded that, to provide timely results for use by Air Force personnel and to maximize findings which will result in reduced training costs for USAF pilot training programs, some version of a pragmatic research strategy appears more useful.

While the final research program for ASUPT will not be required until 1974, planning efforts have been initiated and will be continued. Millions of dollars are spent annually (some \$6-7 billion for all services) for initial and continuation training. While pilots constitute only 2.5% of the total personnel, their training consumes 25% of the total figure. It has been estimated that the savings of one hour in UPT saves from \$750,000 to \$1,000,000 annually. Needless to say, the use of ASUPT capabilities to identify potential savings in pilot training costs will be a significant factor in determining the final experimental design.

References

- Caro, P. W. *Transfer of instrument training and the synthetic flight training system*. NAVTRADEVCE IH-206, Naval Training Device Center, Orlando, Florida, 1972.
- Flexman, R. E., et al. *Evaluation of a contact flight simulator when used in an Air Force primary pilot training program: Part 1: Overall effectiveness*. AFPTRC-TR-54-38, Air Force Personnel and Training Research Center, Lackland AFB, Texas, 1954.
- Gum, D. R. *Development of an advanced training research simulation system*. *Proceedings, Psychology in the Air Force, Third Annual Symposium*, USAF Academy, 1972.
- Kelly, L. L. *The pilot maker*. New York: Grosset & Dunlap, 1970.
- Matheny, W. G., et al. *To move or not to move? The problem of motion in training simulators*. Commemorative Technical Journal, 25th Anniversary, Naval Training Devices Center, Orlando, Florida, 1972.
- Wolff, H. H. *The wide-angle visual simulation problem*. Commemorative Technical Journal, 25th Anniversary, Naval Training Devices Center, Orlando, Florida, 1972.

DEVELOPMENT AND EVALUATION OF TWO FUNCTIONAL PART-TASK TRAINERS

Horace H. Valverde and Bertram W. Cream

Air Force Human Resources Laboratory

This paper describes two projects conducted by the Air Force Human Resources Laboratory in cooperation with the Tactical Air Command. The first project was designed to provide practice in voice communications pertaining to target location for the airborne FAC and strike pilot. Photographic imagery was used as the display source. The second trainer was designed to provide positive transfer of training for tracking and equipment operation tasks for Forward Looking Infrared (FLIR) and Low Light Level TV (LLLTV) sensor operators aboard Gunship Aircraft. TV video tape was used on the display source. Both trainers have been accepted by the users as providing positive training value to their programs.

SECTION I

DEVELOPMENT AND EVALUATION OF A TRAINER FOR FORWARD

LOOKING INFRARED AND LOW LIGHT LEVEL TELEVISION

SENSOR OPERATORS ABOARD "GUNSHIP" AIRCRAFT (AC-130A/E)

Based upon a request from Tactical Air Command Headquarters, Langley AFB, Virginia, and the Gunship Systems Program Office at Wright-Patterson AFB, Ohio, we designed, developed and are currently evaluating a training device to provide positive transfer of training for the tracking and equipment operation tasks performed by Forward Looking Infrared (FLIR) and Low Light Level TV (LLLTV) sensor operators aboard "Gunship" aircraft.

Since the introduction of the first Gunship (the AC-47) "Puff the Magic Dragon") in Southeast Asia, the concept of transport type aircraft armed with side-firing guns has proven of great value.

The first aircraft of this type were fairly simple in terms of the requirements placed upon crew members. Primarily, this was caused by the absence of sophisticated target acquisition systems. It wasn't until the introduction of the AC-119K with a Forward Looking Infrared (FLIR) sensor that a special crew member was designated as responsible solely for this operation. Although the image quality of these early FLIR sensors was poor as compared to today's sets, it was immediately apparent that specialized training was necessary. In the early days of this type of operation, FLIR training was accomplished by a

combination of textbooks, lectures and airborne practice. However, because of the unreliable nature of the equipment, small numbers of available aircraft for training, and the major differences between the practice terrain and targets compared to those in the combat areas, students had to rely upon on-site training to become proficient.

As the new AC-130A/E Gunships entered the inventory, it became obvious that some sort of ground training device was required to enable the sensor operators to practice their tasks without the requirement of actual flight. It was to fill their need that the trainer was designed.

Design of the Trainer

Preliminary Trainer Considerations

Because of the cost and difficulty in obtaining actual equipment, it was decided that use of actual infrared or low light level TV equipment in the trainer was impractical. Because of this, it was decided to concentrate upon providing faithfully reproduced imagery from these sensors. Naturally, it was also our aim to make the trainer reliable and easy to operate. As in other projects of this type, the emphasis was placed on psychological rather than engineering simulation.

Psychological simulation concentrates on those particular aspects of a task that are both critical to job performance and provide positive transfer of training. Engineering simulation requires a one-to-one duplication of actual equipment and consequently drives the cost of any part-task device beyond that which can be considered economical. Because psychological simulation places great reliance upon the accurate identification of training objectives, it is necessary to have complete knowledge of the tasks that will be performed in the combat environment.

It was understood from the beginning of this project that the device would be prototypical in design so that information gained might contribute to future efforts of this type. The intention then was two-fold. First, to provide a useful training device that would fill a stated field requirement, and second, to serve as a test-bed for application of new training techniques that might, in the future, be applied to other advanced systems.

17

Description of Multisensor Trainer

The trainer uses an active CRT type display for image presentation with video tape as the image source. It also includes the capability to expand field of view of the imagery in the same scale as that in the actual equipment. (For a full description of the operation of the video distribution system see Appendix). The imagery may be actual FLIR or LLLTV taken from the onboard sensors, or a video tape of a rotating terrain model with targets.

The primary tracking control panel of the trainer is the gimbal control and joy stick. This panel, which is a duplicate of that on the aircraft, incorporates both a drift and sensitivity control as well as other necessary lights and switches. The "joy stick" has a button on top that serves as a slewing control. This button electronically slews the imagery in both azimuth and elevation in a manner highly similar to the actual equipment. It is this image slewing capability that enables the student to practice tracking. A reticle is etched on the face of the CRT. The task is to identify a potential target in the display and then, by use of the slew button, center and maintain the target in the reticle regardless of image motion. Because the gimbal panel also contains a drift control, (which compensates for gimbal drift in azimuth and elevation) the student is able to refine his tracking skills. This is done by using the drift control as the only means of centering and maintaining the target in the reticle.

We felt that it was also desirable to provide the trainee with as many accessory equipment panels as possible. First, to familiarize him with approximate locations of the various components; and second, to enable practice of equipment operation procedures. This also allows the instructor to explain these procedures with visual aids and also allows insertion of system malfunctions. Toward this end, we included the following associated equipment panels.

a. *FLIR Control Panel.* This panel, along with the gimbal panel, controls the operation of the FLIR sensor. We included an operating search/track switch (which controls the field of view of the sensor optics) as well as having the system enabled by the operate select switch.

b. *The Sensor Angle Display* panel is a working mockup. Its purpose is to indicate to the sensor operator the position of his sensor head in relation to the other sensors. On our trainer, the indicator needles may be positioned by the instructor to provide graphic demonstration of the equipment operation. In addition, it may be used to present problems to the students when discussing firing geometry and target/sensor orientation.

c. In keeping with the distinction previously made between engineering simulation and psychological simulation, we decided to represent the 28VDC circuit breaker panel and control switch unit panel by engraved plastic representations. Although neither operates, they are useful in providing the student the opportunity to see their placement and be taught their function in relation to the other equipment at his station.

d. One addition that was made which is not part of the actual equipment is the image recenter button. Due to the type of display used, it is possible for a student who has maximum gain set on the drift

control to actually slew the image so that it leaves the visible portion of the CRT. If this occurs, and if he cannot readily recenter the image, depressing the button will do the recentering for him.

e. The Remote Control Unit and the Intercom Unit both contain functional panel lights and follow the rationale of psychological simulation. Both of these units serve as procedure training devices.

Evaluation

The multisensor operator trainer is being evaluated by both the students and the instructors. The two major factors being measured are degree of tracking accuracy and equipment operation ability. Seven classes were selected. Each class of sensor students was divided into two groups, control and experimental. The experimental group received training on the new device while the control group received only their normal training. Performance was evaluated during the 11 flights that comprise the airborne training for all sensor operators.

Results

Preliminary data from the ratings of student tracking and surveys of the instructors and students indicate that the trainer is performing as designed. Also, it appears that the trainer enables the new student to reach his tracking asymptote earlier than was previously possible.

Using Chi-square, the experimental groups show a significantly greater ability to correctly perform the required tasks than the control groups. The difference between correct and incorrect performance percentages are greatest during the first seven missions. After this point, both groups show approximately the same performance level. As we can see from Figure 1, the experimental group shows accelerated skill acquisition curves for both the equipment preflight and target tracking behaviors.

Using the Mann-Whitney U Test, the difference between the number of missions required by the control and experiment groups to reach the criterion of a 3 rating is significant at the .001 level for tracking and at .015 for equipment preflight. In addition to allowing ground based practice of the required skills, it enables the slower student to receive remedial instruction without the requirement of actual flight. The users of this device are enthusiastic about its training value, and have incorporated it into their formal curriculum. Because it allows the student to reach his required skill level sooner, the remaining flight time may be devoted to the practice of more sophisticated skills, such as image interpretation.

As a result of the success to date, TAC has requested our assistance in applying the technology gained from this project to the design of similar devices for the other Gunship flight crew stations.

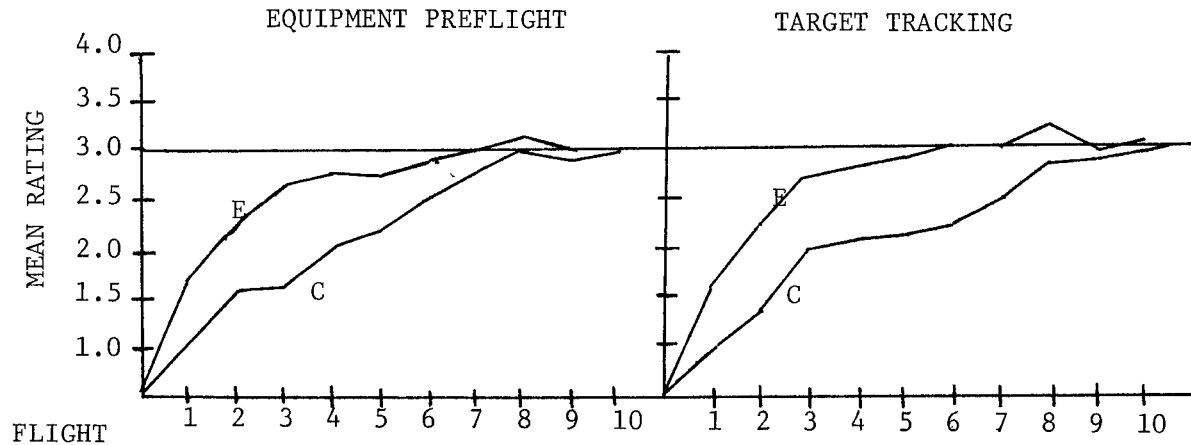


Fig. 1. Skill acquisition curves for equipment preflight and target tracking behaviors.

Conclusions

The multisensor operator trainer has demonstrated the feasibility and utility of functional part-task trainers. Central to this is the emphasis on psychological simulation for task relevant behaviors.

The information gained from this project will be applied to the design of training devices for application in advanced Air Force systems of the future.

SECTION II

DEVELOPMENT AND EVALUATION OF A PART-TASK TRAINER

FOR COMMUNICATING TARGET LOCATIONS

This section describes the development and evaluation of a part task trainer designed to help the forward air controller (FAC) become more proficient in communicating target locations to the tactical strike pilot. For the purpose of brevity, we will refer to the strike pilot as the TAC. The FAC/TAC ground trainer was developed to permit low flying FACs to practice communicating target locations to high flying TACs.

In limited war/counterinsurgency operations, the FAC and TAC engage in airstrikes in (a) close support of ground forces and (b) those not in support of ground force activities. Also, the FAC performs visual reconnaissance missions which do not include airstrikes. The FAC/TAC trainer was developed primarily to provide practice in voice

communication of target locations. During strike operations, pilots of tactical aircraft are directed by FAC pilots who operate relatively slow, low flying aircraft such as the O-2 and OV-10. In this type of operation, the FAC visually identifies targets and directs the TAC to them by voice communications. Thus, in one sense, the FAC serves as the eyes of the TAC. The FAC also must be able to describe target areas in terms of landmarks which are distinguishable by the TAC flying at higher altitudes and faster airspeeds.

One of the most difficult aspects of a FAC's duties is to make certain that the strike leader sees the target or target area. Many techniques can be used to achieve this objective. The easiest way is to mark the target from an airborne platform. Normally, the fighter aircraft will have the FAC in sight and watch him roll in to fire a rocket which will mark the target. However, many circumstances might preclude the FAC from marking and he must verbally describe the target. Some of these circumstances might be: (a) heavy ground fire (which the FAC cannot survive), (b) the desire for complete surprise, (c) the FAC has no marking rounds left, and (d) the FAC is controlling from a ground position. Whether the FAC can mark a target or not, he must be able to communicate its location to the higher flying TAC. When the strike pilot has the target area in sight, the control of the strike is very simple. The FAC needs only to clear each fighter as they attack the target. The FAC/TAC trainer provides practice in target location communications to improve FAC/TAC operations.

Method

Trainer Description and Operation

The trainer includes positions for the FAC, TAC and instructor. It is constructed of lightweight aluminum tubing and rear projection screens. Other equipment includes a cassette tape recorder, two 35mm slide projectors, an audio amplifier, headsets, and a set of 35mm slides. Also, the three participants in the training session, (FAC, TAC and instructor) share an intercom system and all communications are tape recorded by the instructor. Playback of the recordings is useful in informing the student as to the effectiveness of his description of target locations.

The target area, depicted on slides, is projected on the back of the FAC's and TAC's screen. The instructor uses a pointer, or a pen light, to indicate on the back of the FAC's screen the target to be described by the FAC to the TAC. The instructor turns on the tape recorder. The FAC then establishes contact with the TAC and begins transmitting target information to him. A sample tape recording of a FAC/TAC communication scenario is included with the trainer for demonstration purposes.

When power is on, both students and the instructor can speak to and hear each other through the headsets. If a student wishes to speak privately with the instructor, he presses his pushbutton which lights the respective red light on the intercom set. The instructor pulls the toggle switch corresponding to that student and a two-way conversation is possible. When the toggle switch is released, three-way communications are reinstated.

Trainer Test Imagery

Photographic imagery which compensates for differences in FAC and TAC altitudes was used in the test. The altitudes as represented in the 35mm slides are 2,000 feet for the FAC's screen and 8,000 feet for the TAC's screen. Each scene for both FAC and TAC was photographed from east, north, west, and south. The scenes were photographed in color from a Cessna 172 aircraft at 8,000 feet altitude. Two cameras, mounted side-by-side, were used to obtain the imagery. One camera, equipped with 50mm lens was used to photograph the TAC imagery (8,000 feet). Another camera, equipped with 200mm lens, was used to photograph the FAC imagery (2,000 feet, simulated). Both cameras were operated simultaneously in photographing each scene.

Subjects

A total of 35 Air Force pilots, selected to be trained as airborne forward air controllers, participated in the comparison study. FAC trainees received ground training at Hurlburt Field in O-2 or OV-10 flying training at Holly Field. Both fields are located within the Eglin Air Force Base complex in Florida.

The subjects ($N = 35$) were divided into two groups: Group A ($N = 18$) and Group B ($N = 17$). Group A subjects were pretested and posttested on the trainer which amounted to approximately two hours (one hour for each test). In addition, they received two hours of supervised practice on the trainer. Group B subjects were pretested and posttested, but received no additional practice on the trainer. The average flying time for the two groups was 651 hours. Because of conflicts in trainee entry and scheduling problems, it was necessary to randomly assign subjects for the experimental treatment. However, as a group, they proved to be homogeneous in terms of aptitude and experience. None previously had received specific training in the communication of target locations between forward air controllers and strike pilots. Therefore, differences in the performance of the two groups on the criterion test may be attributed to training experience.

Program Administration

The criterion test consisted of two sets of 10 color slides; one set for the FAC with a simulated altitude of 2,000 feet and another

set of 10 slides for the TAC taken at 8,000 feet. The same number of slides of an equal order of difficulty were used for the pretest and practice session.

The testing program was administered during scheduled class periods. Group A and Group B were pretested and posttested on an individual basis. Two experimental subjects received supervised practice on the trainer at the same time. They alternated roles; as FAC for one hour and as TAC for one hour. Thus, Group A performed ten trials for about one hour during the pretest; twenty trials for about two hours during the supervised practice session; and ten trials for about one hour during the posttest. Group B performed ten trials each for pretest and posttest. The subjects were scored on the basis of time and accuracy in communicating target locations.

After completing the posttest, each subject was administered a questionnaire. The questionnaire was used to assess the subject's general attitude toward the trainer.

Results

One of the first considerations in the design of this experiment was to determine the gain in proficiency a student might achieve through practice on the trainer. Ratio gain scores were calculated for each subject for the pretest and posttest. The ratio gain scores are obtained by dividing the actual gain by the possible gain. The results can be used for comparison purposes.

A Mann-Whitney U test was calculated for the rank-order data to evaluate the difference in the achievement in gain between the two groups. The gain of Group A was significantly greater than that of Group B (Mann-Whitney U, $z = 2.475$, $p > .01$, $< .05$).

Supervised Practice

As indicated previously, during the practice session the Group A students performed on the trainer in teams (alternating FAC and TAC roles). The students' performance as teams was recorded by the experimenter. The students were provided immediate feedback as to the results of their efforts. Scoring of the targets provided structure for the practice sessions as well as knowledge of results for the students.

Trainer Evaluation Questionnaire

The first 8 items on the 11-item questionnaire were intended to provide an indication of student satisfaction with the trainer. The remaining 3 items were concerned with how the student felt about (a) the difficulty of the training, (b) length of time spent in training, and (c) how much of the same training was previously received by the

student. Students responding in the top 40% of the scale (favorable end) were given a score of one and those responding below the top 40% were given a score of zero. A total "general satisfaction" or "favorable attitude" score was obtained for each group of students from the first 8 items of the 11 item scale. The percentage of students responding to the two items on the favorable side of the scale was as follows: (a) experimental = 80.6%, (b) control = 80.1%. However, of the total number of subjects (35), only 3 gave a negative evaluation of the trainer.

Concerning the last 3 items (9-11), most of the students (27) expressed the opinion that the training content was not difficult. Although the testing varied from 4 hours for Group A to 2 hours for Group B, 10 Group A and 9 Group B students felt that the time spent in training was adequate. Additional training time was felt needed by 8 Group A students and 6 Group B students. A total of 32 students responded that for them the training content was almost entirely new material, or that it only slightly overlapped with other training received.

Conclusions

Initial performance by the subjects showed a wide variance in speed for single trials as well as for entire test sessions. There were frequent reversals of thought and confusions over landmarks, directions, visual cues, and misinterpretations of apparently straightforward instructions. A few FAC subjects evidenced a unique ability to start with easily seen, gross landmarks and then lead the TAC listener to the target by clearly specifying figure/ground relationships in the visual scene with a minimum number of statements. This ability may be the essence of the verbalization of imagery task and its optimization worthy of training.

The trainer can be used to teach communication of target locations which seems to require an unusual amount of subjective judgment, experience, or native ability. Entire missions can be accomplished in view and hearing of a class because of the elevated location and large size of the photographic imagery and the addition of a loud-speaker in the intercom system. The device also can be used for remedial instruction in target detection, recognition, and identification.

Based on the results of the analysis of individual trials, the instructor may opt to provide FAC students with 2 to 4 hours of practice on the trainer. The additional 2 hours of practice by Group A subjects permitted them to reach asymptote after the first trial of the posttest. Group B subjects who performed only 2 hours on the trainer, had not reached their asymptotic level by the tenth trial of the posttest. These results may assist training personnel in making training media utilization decisions. For example, it may be

necessary to make a tradeoff between a high proficiency level and training costs. In the present case, the results indicate that for optimum skills acquisition, 4 hours of practice should be provided for communication of target locations.

This has been a brief overview of two related projects in which part-task trainers have been used to enhance the practice of skills identified as having positive-transfer to the operational situation. Both have served to validate the concept of functional training devices whose emphasis rests on psychological rather than engineering simulation.

The results of the evaluations made will be applied as appropriate to the design and development of training programs and devices for advanced systems of the future.

Human Factors (Dual Session)

CHAIRMAN: Captain Frederick V. Malmstrom

EFFECTS OF ELECTRONICALLY PRODUCED AIRBORNE NOISE ON
PSYCHOPHYSICAL PERFORMANCE OF MILITARY TASKS

James P. Jenkins

Naval Ship Systems Command

The U. S. Navy has participated in a number of studies with emphasis placed on noise effects as associated with military mission effectiveness. A series of experiments were conducted on the psychophysical effects of short duration (250 msec to 1 sec), low frequency pulses almost continuously produced by electronic equipment. These pulses were presented for 15 twenty-four hour days and 30 twenty-four hour days to 20 subjects during two experimental sessions. Sound pressure levels of 85 dB - 90 dB were tolerated without deleterious effects on hearing, sleeping and military type performance tasks. Social psychological data were obtained and stimulus effects were minimal.

The military departments in the Department of Defense have been concerned with the physical and psychological effects of noise on its personnel and on the civilian populace. This paper will describe the most recent study performed by the Naval Ship Systems Command, Department of the Navy, to determine the effects of electronically produced airborne noise on the physical and psychological behavior of personnel performing a wide variety of military tasks over an extended period of 30 days. The noise was in the form of simulated sonar transmissions. This study was specifically directed toward certain classified problem areas and detailed reports are available to those with proper clearance and need-to-know. For this reason, some aspects of this program are classified and are omitted herein.

The objective of the study was to:

- (a) Determine the effects of electronic transmissions on hearing and on performance in typical tasks, such as vigilance, tracking, reaction time problem solving and computational activities; to determine the effects on psychological reactions and group behavior and such effects as may be related to quality and quantity of sleep.
- (b) Verify previously established sound pressure levels which would not result in unacceptable performance.
- (c) Extend the results, where possible, to other situations having related acoustical environments.
- (d) Provide the Navy and other military departments those facilities and a larger data base for additional research, as required.

Performance tasks. Prior studies (Kryter, 1970) have reported both the auditory effects and effects on work performance due to the internal disruption of the perceptual processes by noise. This may be contrasted with a number of other studies on noise resulting in irritability, annoyance, speech interference and the like. Before proceeding further we wish to define what tasks may properly be called military tasks, as compared with other tasks which may be common to a variety of work situations. Military tasks are behavioral tasks called upon in a variety of military situations and are those parts of the man-system interface which are required to fulfill a military mission requirement. Because these tasks are usually associated with a system they are likely to be complex. For this reason we would exclude purely motor tasks such as, for example, hauling a hawser, carrying a parcel, digging a foxhole or similar tasks. In themselves they are only secondarily related with a system, although performing them may be antecedent to mission success. The military tasks defined in this paper were those performed in complex man-machine system, such as a ship, aircraft, or submarine. They were related to the sensory-decision making-control paradigm familiar community; the major tasks were:

- (a) Compensatory tracking
- (b) Visual attention
- (c) Auditory vigilance
- (d) Visual Reaction Time
- (e) Memory
- (f) Mental problem solving
- (g) Speech intelligibility
- (h) Visual vigilance

Certain other tests were given as well, but will not be described in this paper.

Literature review. In general, it has been found that noise does not have deleterious effects on mental or motor task performance, although Broadbent (1954) offers the theory of blinks, which postulates that stimulation of the auditory system in turn affects the central nervous system and disrupts sensory perceptions in a manner analogous to the blinking of the eye. Several researchers have applied Broadbent's theory with mixed results. C. S. Harris (1968), offered the idea that the vestibular system is next affected after the auditory in the presence of noise levels greater than 120 dB which, when presented, both symmetrically and asymmetrically, resulted in increased error in psychomotor tasks. Test results are open to considerable explanation which do not fully support the concept of vestibular involvement. Our levels were well below 120 dB.

This brief review of the literature leaves one with a sense of dissatisfaction and poses several interesting hypotheses. If, as in the case of many military environments, an ambient background noise

below 75 dB (A) re 20 μ Pa has superimposed upon it at certain intervals tones clearly audible for the frequencies and band involved, would there result performance decrements affecting satisfactory military task completion. A null hypothesis was postulated. Performance of military tasks are not affected by electronic transmission in the form of simulated sonar pulse tones over considerable periods of time. In the sections which follow two experiments, their results and analyses are presented.

Method

Two experiments were performed in the laboratory and a third, yet to be conducted *in situ*, is planned. Subjects for both experiments were from the same population. They consisted of 20 male, Naval enlistees who volunteered for each experiment. All passed medical and audiometric examinations and had no observable defect physically or audiometrically. Their ages were from 18 to 34 with a mean of 19.9 years. Their military experience spanned from a few months of service to over 16 years, but the majority had about 16 months in the Navy. Ranks of the subjects reflected their time in service and ranged from E-3 to E-8, again with most having attained an E-4 rank, which is commensurate with their length of service.

Stimulus

The stimulus condition for both experiments were similar, but not identical. With an ambient background level of 55 to 65 dB (A) re 20 μ Pa, pulsed tones of simulated sonar transmissions were presented every minute or less, continuously 24 hours per day for a minimum of 15 days in Experiment 1 and a maximum of 30 days in Experiment 2. In Experiment 2, 15 days of pretests and 10 days of posttest performance measures were obtained with a normal ambient background level. Center frequencies were between 3.0 and 4.0 kHz and duration of each pulse was from several hundred milliseconds to slightly over two seconds. Levels were systematically varied from 80 to 85 dB in Experiment 1 and from 80 to 90 dB in Experiment 2. Thus, the aural stimuli were highly similar to those conditions resulting from certain sonar used by the Navy.

Procedure

The core of the performance test facility was the LINC-8 Computer (Digital Equipment Corp.). This machine is specifically designed for controlling laboratory experiments and has flexible provision for stimulus presentation, event control, response acquisition, data tabulation, and output. In general, stimulus displays were generated on the LINC-8's 5-inch CRT and transmitted by a closed-circuit TV system to the test subjects in the Performance Test Room.

The Performance Test Room served as the major testing area. The S sat at a desk-chair and responded to displays shown on his TV monitor (Setchell Carlson Model 10M915). A general purpose response panel was mounted on each S's desk, and cabling connected these panels to the LINC-8 to provide for response scoring. Four pushbuttons were fixed to its upper surface, and a single button was located at the left hand side of the panel's forward edge. A control stick was mounted at the right hand side of the panel. A phone jack receptacle made it possible to plug S's headset into the panel for auditory detection tasks. Each subject underwent tests in the morning and afternoon. The group of 20 Ss was divided into Alpha Beta teams of 10 Ss each. Table 1 depicts a typical test day schedule. Alternate test schedules were used so that each day's activities were not the same. Each S had about five hours of test per day.

Tracking. This experiment tapped the general class of manual control skills required in the exercise of ship control. Each S worked on a compensatory tracking problem in which his control stick moved a spot of light on his TV monitor. The task for S was to keep the spot centered (from left to right) while acceleration control dynamics residing in the LINC-8 computer served to complicate the tracking problem. Ten Ss were run at once. The LINC-8 sampled each S's control stick and displayed the 10 spots on its CRT display. A closed-circuit TV system routed this display to ten TV monitors, each of which was physically masked to permit viewing of only a single spot by each S. The acceleration parameter was adjusted over the days of testing to provide three levels of problem difficulty.

Visual attention. In this task the S performed a logical test on a random triplet of digits briefly presented on his TV monitor. If the digits satisfied the test, the S was to respond by pressing a button before the display was blanked. The logical test was: Press the button if (a) the 1st digit is largest and the 2nd digit is smallest; or (b) the 1st digit is smallest and the 3rd digit is largest. Given lengthy presentation times this is quite an easy problem; as exposure times are decreased the task provides a useful measure of attention or alertness. Exposure times were reduced from 3.42 sec. to 0.87 sec. over the course of the experiment. The number of daily trials varied from 350 to 700 in order to maintain a test of approximately 20 minutes' duration.

Visual reaction time. In this task four letters of the alphabet were displayed in a row on S's TV monitor and at random times (from 3 to 18 seconds), one of the letters changed to a new one. The S was instructed to press that one of four buttons on his response panel corresponding to the position of the letter which changed. One hundred trials comprised each daily session.

Memory. A recognition-memory task was given every other day of testing. Here S was exposed to 100 words, presented one at a time

TABLE 1
Typical Test Day Schedule

TIME	GROUP ALPHA	GROUP BETA
0700		
0800	BREAKFAST H2	H2
0900	H8	T
1000	T	H8
	/	A
1100	A (E)	VA
1200	LUNCH	
1300	VA	P
1400	/	RT
	RT	/
1500	+	+
	/	I
1600	I	/
	ψ	ψ
1700	/	I
	I	/
1800	/	/
1900	DINNER	

H2 = Hearing tests (2 subjects)
 H8 = Hearing tests (8 subjects)
 T = Tracking
 A = Attention
 VA = Vigilance - auditory
 E = Exercise period (optional)
 /// = Free time
 P = Problem solving
 RT = Reaction time
 + = Addition
 M = Memory
 S = Sonar
 I = Intelligibility
 ψ = Psychological testing

on his TV monitor. These constituted the words to be learned. The recognition phase began after a delay of 30 seconds. At this time the S was handed a list of 200 words comprised of the 100 "TV words" and 100 additional words, all in random order. The S gave each word an integer rating from 1 to 4, the higher ratings indicating greater confidence that the given word *was* in the TV presentation. Analysis was in terms of the Theory of Signal Detectability.

Problem solving. A variety of paper-and-pencil problem-solving tasks were administered each day. These included tests of reading comprehension, arithmetic reasoning, maze and path tracking, pattern recognition, etc.

Results

Experiment 1

In this experiment we found a steady learning period, during transmissions, flattening to a plateau for all performance tests. Sonar transmissions did not seem to affect performance, but because baseline performance data was not available prior to start of transmissions the conclusion remained to be verified, as described in Experiment 2. This experiment should be looked upon as obtaining classic learning curves.

Experiment 2

In this test performance measures prior to transmissions produced more varied results and incomplete learning. After transmissions began reaction time performance exhibited a decrease ($p < .01$) compared with baseline, but operationally the difference does not seem meaningful, e.g., 91% to 87% correct (See Figure 1). Reaction time values are shown in Figure 2, although mean reaction decreased slightly the results are insufficient to account for the larger decrease in percent correct. All other performance tests results were constant throughout the transmission period and during post test periods and effects of electronic transmissions were not significant.

Discussion

It may seem that the null hypothesis of electronic transmissions up to 90 dB (A) not affecting performance of military tasks was not completely rejected. Why was reaction time performance significantly different during and after exposure than before exposure, whereas performance results of memory, visual attention, arithmetic reasoning and the like did not reveal differences? Research results from unpublished Navy studies similar to this do not substantiate differential effects on performance due to sonar transmissions. However, identical tests to the ones given in this project were not included.

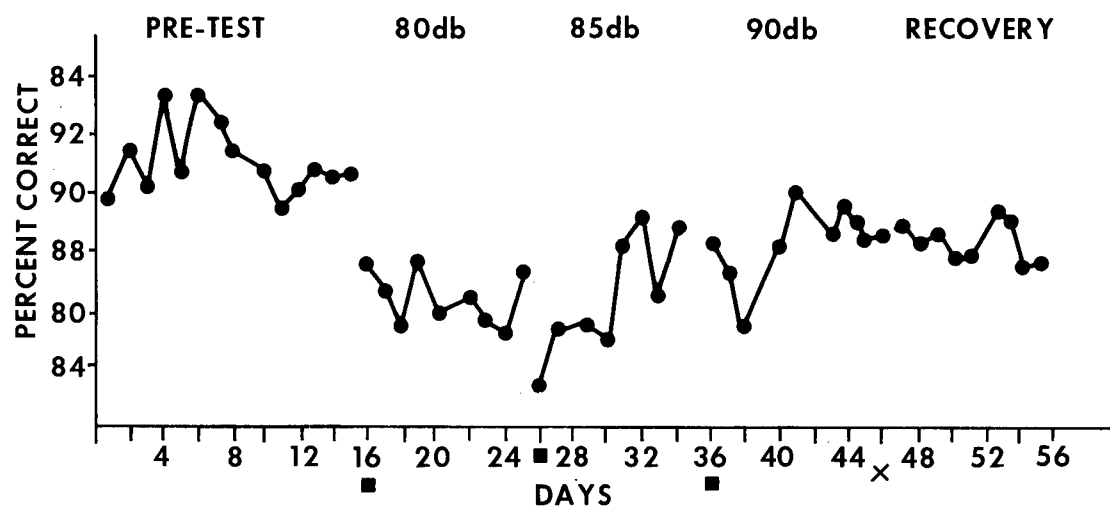


FIGURE 1

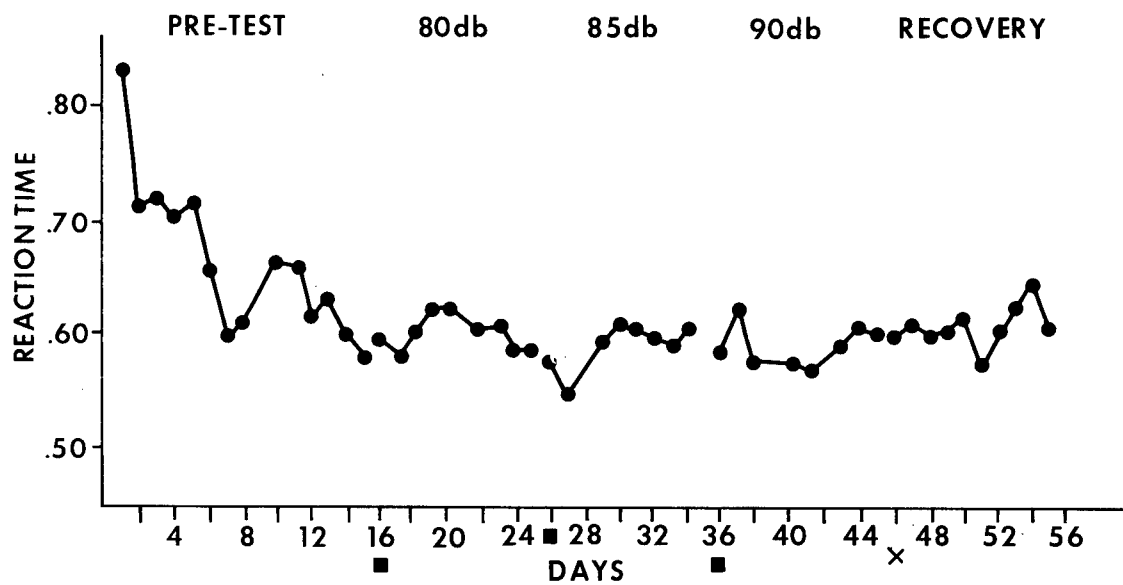


FIGURE 2

The findings on reaction time do not seem to fit conclusions reached by Kryter in his review of effects of noise on mental and motor performance. If the transmission effects were related to type of tasks, perceptual-motor, as say compared with cognitive only, then other tasks such as compensatory tracking or visual attention could be expected to also demonstrate negative changes. The same conclusion could be drawn if one applied reinforcement theory inasmuch as subjects received feedback immediately after each test session. An information content analysis was done to determine whether the bits per second rate was changed and the findings were that they were not. Further, no recovery occurred after transmissions ceased. And finally, although performance decreased with onset of transmissions, an increase in the level from 80 dB (A) to 85 dB (A) or to 90 dB (A) did not relate to a step function decrease in performance. Consequently, we reject the explanation that the transmissions or other independent variables under control contributed to decreased performance. We do not offer a viable alternative explanation.

However, the results from all tests give strong support to the position that a wide range of military tasks can be satisfactorily performed in the presence of noise, similar to that used here, without serious performance decrements. An alternative explanation may be offered: the performances of reaction time tests were attributed to learning and a stable performance was obtained after the first 16 days and did not change thereafter. Transmissions had no effect. This explanation does not follow other test results, i.e., no significant differences, nor the results of the auditory vigilance test. An adequate explanation cannot be offered now. Some additional tests of performance at sea are planned and these results may provide better understanding of the effects of electronically produced airborne noise on psychophysical performance of military tasks.

References

- Broadbent, D. E. Some effects of noise on visual performance.
Quarterly Journal of Experimental Psychology, 1954, 6, 1-5.
- Harris, C. S. *The effects of high intensity noise on human performance*.
AMRL-TR-67-119, Air Force Systems Command, Wright-Patterson AFB,
Ohio, 1968.
- Kryter, K. D. *The effects of noise on man*. New York: Academic Press,
1970.

Acknowledgment

This project was under the overall direction of the Naval Ship Systems Command, with technical direction by Naval Undersea Research and Development Center. Mr. Robert Gales was Principal Investigator and Mr. Ray Hershman, Naval Electronics Laboratory Center, Senior Scientist for Performance Testing.

"CAN YOU REACH THE CONTROLS?"

A COCKPIT ANTHROPOMETRIC SURVEY

Harvey G. Gregoire

Naval Air Test Center

Unique equipment and improved procedures developed at the Naval Air Test Center have proven accurate in identifying and quantifying cockpit dimensions which will cause particular aviators to be less effective or compromise safety of flight as a function of body size. Results of the survey revealed two aircraft with controls beyond the reach of even the 95th percentile pilots and five aircraft with controls beyond the reach of the 50th percentile pilots. In the procurement of future airplanes, strong emphasis must be placed on designing the location of all controls to be within reach of the specified anthropometric range of aviators.

The inability to sit comfortably in cockpits or easily reach controls requiring immediate actuation, continues to be a significant problem in certain airplanes due to anthropometric incompatibilities with particular pilots. New equipment and improved procedures have proven accurate in identifying and quantifying cockpit dimensions which will cause particular aviators to be less effective or compromise safety of flight as a function of body size. Pilot-cockpit compatibility should be one of the criteria in the assignment of aviators to particular airplanes in the present inventory. In the procurement of future airplanes, strong emphasis must be placed on designing the location of all controls to be within reach of the specified anthropometric range of aviators.

Studies of accident and incident records at the Naval Safety Center have revealed that pilot anthropometric incompatibility continues to be a significant problem in safe and efficient aircraft operation (see Reference 1). Unique equipment and improved procedures have been developed at the Naval Air Test Center to quantify functional reach distances to essential controls. Functional reach varies according to body size and seat height adjustment. The Office of Naval Research requested that specified aircraft be surveyed to determine reach distances to emergency controls in an effort to provide data which could be used to identify the degree to which particular aircraft cockpits were deficient in anthropometric compatibility.

A survey of typical Navy tactical airplanes was accomplished to acquire the following data:

a. Excessive reach-to-emergency control distances as related to the 5th, 50th, and 95th percentile aviator dimensions specified (see References 2 and 3).

b. Available sitting height.

c. Obstructions to control reach.

The data gathered in the survey have potential for use in establishing an aviator assignability code which could be used in assigning aviators to aircraft which would be compatible with the particular aviator's physical dimensions such as sitting height, functional reach, etc.

Method

The airplanes surveyed in this study included the A-4C, A-6A, A-7E, AV-8A, F-4J, F-8D and OV-10A. The data for a particular model airplane cannot necessarily be extrapolated as applicable to a similar series aircraft, e.g., the A-4C data may not be identical to A-4M data, A-6A data may not be the same as A-6B, etc.

The measuring equipment consisted of a G-2 Anthropometer. The anthropometer is adjustable in three dimensions to simulate all percentile ranks of sitting shoulder height, sitting eye height, eye depth, and bideltoid diameter. Photographs were taken to illustrate some of the actual pilot reach-to-control incompatibilities and to verify anthropometer measurements.

The data acquired consisted of functional reach distances required of 5th, 50th, and 95th percentile aviators to reach emergency controls. The measurements taken were based on criteria that assumed an aviator was positioned in the seat with his back held securely against the seat back by a locked shoulder harness. Some of the most critical instances of reaching for emergency controls occur when the aviator's back is held firmly against the seat by g forces or by the locked shoulder harness, such as during catapult launch, "bolter", weapons delivery, landing, "jinking", etc. The aviator-back-to-seat juxtaposition used for the cockpit measurements approximates the relationship used in determining functional reach when measuring pilots' anthropometric dimensions.

In the data collected, allowances were made for the Nomex summer flight suit, MA-2 torso harness, and MK-3C flotation device. Correction for factors such as slouch, sag, or stretch was not applied to the data due to the wide variance in such factors. In each airplane evaluated, a "maximum reach" survey was also accomplished to establish how much farther beyond normal reach an aviator could stretch by maximum exertion. Maximum reach varied as a function of the design characteristics of restraint systems.

Total sitting height available was measured in each airplane. Obstructions and inadequate access which hampered safe and rapid actuation of emergency controls were identified during the evaluations.

The basic procedure used in obtaining the data is described below. Although 5th percentile anthropometric dimensions are specified in the description of procedure, the steps were repeated for 50th and 95th percentile dimensions. The procedure was:

- a. Anthropometer was adjusted to the sitting eye height, eye depth, sitting shoulder height, and shoulder breadth (bideltoid diameter) of a 5th percentile aviator.
- b. Anthropometer was placed in the seat in such a manner that the rear surface of the anthropometer was in contact with the forward surface of the seat back.
- c. The seat was adjusted to place the 5th percentile eye location at the cockpit Design Eye Position (DEP).
- d. With the seat adjusted and the anthropometer in place, the retractable measure was extended from the respective shoulder point to the controls, and the distances read directly from the retractable measure.
- e. Distances to controls were recorded and corrected for angles of azimuth and declination from the shoulder back tangent point to the particular control.
- f. Total sitting height was measured from the compressed seat cushion to the closed canopy surface at the point where the aviator's head would contact the canopy if there were insufficient sitting height. Sitting height was measured with the seats adjusted FULL UP and FULL DOWN.

Measured distances to controls requiring functional reach, i.e., grasp, were compared to the functional reach capabilities of 5th, 50th, and 95th percentile aviators listed in Reference 3. Reference 3 was utilized in the latter case because of the lack of fingertip reach data in Reference 2.

"Maximum possible reach" data were collected in the following manner for each airplane: an individual with 50th percentile arm reach, 50th percentile sitting shoulder height, and 70th percentile bideltoid diameter was attired in a Nomex summer flight suit, MA-2 torso harness, and MK-3C flotation garment. The subject firmly secured himself into the ejection seat with lap belt and torso harness, and pressed his back into the seat to simulate the effect of g force during catapult launch, "bolter", "jinking", etc. The subject first

reached for controls in a normal fashion without stretching; measurements were taken from the subject's fingertips to the controls. The subject then reached for the same controls exerting forward force against the harness and stretching as much as possible; measurements to the controls were repeated, and subtracted from the prior normal reach. The differential quantity represented a stretch beyond that which is comfortable or at times even possible in dynamic airplane operation.

The reference for reach differential measurements was a line extending directly forward from the shoulder of the arm reaching for the controls. Controls in the reach area for the reach differential measurements were generally 0 to 15° in azimuth and 0 to 35° in declination from the shoulder reference point. Each measurement of reach was repeated five times for each hand in each airplane. The differences were then averaged to quantify the reach differential between normal and maximum possible stretch.

Special instrumentation included the locally designed and fabricated Model G-2 Anthropometer.

Cockpit evaluations are performed in every stage of an aircraft weapons systems development. These evaluations include analytical methods employed to investigate mission, tasks, functions, operations, work load, design layouts, etc., and progress to computerized mathematical models and engineering mockups. Finally cockpit evaluations are made in prototype mockups, and in complete operational cockpits. In spite of this evaluation program, the Navy continues to have aircraft flying in the fleet which are not anthropometrically compatible with the aircrew population, and in some instances are not safe for the operational user.

20

There are many causes for the failure to detect anthropometric problems during the aircraft weapons systems development. Among the causes are (a) lack of emphasis on human engineering design considerations necessary to accommodate the entire anthropometric range of the potential operator population, (b) lack of adequate specification definition and enforcement in the contractual procurement of airplanes, and (c) waivers to specifications during airplane procurement which are granted by nontechnical administrative personnel who are not in a position to understand fully the effects of granting such waivers.

To achieve maximum effectiveness and safety, aviators should be assigned to present inventory of airplanes with the least amount of cockpit/anthropometric incompatibility relative to their own anthropometric dimensions.

The distances beyond the reach of 5th, 50th, and 95th percentile individuals to the particular controls are presented to an accuracy of tenths of an inch. One decimal place is used due to the accuracy

of the anthropometric data in References 2 and 3, not because of one decimal place accuracy of the measurement equipment or procedure.

The amount of vertical seat adjustment is noted in each airplane data table.

Photographs of a subject reaching for particular controls in each airplane were taken to verify data derived from the anthropometric measurements. The subject was attired in a Nomex summer weight flight suit, MA-2 torso harness, and MK-3C flotation garment. The subject's physical dimensions were: functional reach 30.3 in. (10th percentile), shoulder height 22.5 in. (25th percentile), sitting height 34.5 in. (15th percentile), sitting eye height 30.3 in. (15th percentile), and shoulder breadth 17.1 in. (20th percentile). The subject selected for photographic documentation was used to illustrate the more common reach deficiencies which hamper the smaller anthropometric range of aviators, e.g., 25th percentile.

Results

The data acquired are presented in Tables 1 through 7 for the seven airplanes evaluated. In Tables 1 through 7 a dash (-) indicates that the control was within the reach of the particular aviator size surveyed. An asterisk(*) preceding particular controls identifies those controls judged to be critical by aviators who had recently completed operational combat tours or extensive test flying in each of the airplanes evaluated. The controls were determined to be critical if they were likely to be actuated in an emergency during which the aviator was held against the seat by a locked shoulder harness.

It should be noted that the data demonstrated that functional reach and fingertip reach deficiencies are more numerous and more extreme for the average and smaller sized aviators (50th percentile and smaller). This is a result of the fact that the ejection seat rails slant aft so that upward seat travel to attain DEP moves the pilot farther from the controls.

Conclusions

The data acquired in the cockpit anthropometric survey demonstrate various degrees of anthropometric incompatibilities in each of the cockpits evaluated. The most serious incompatibilities were critical controls beyond the reach of the entire range of aviators' functional reach in the A-7E (Table 3) and the OV-10A airplanes (Table 7).

The functional reach and fingertip reach deficiencies were more numerous and more extreme for the average and smaller aviators (50th percentile and smaller), who, because of the seat back angle, are forced to travel backward and upward in order to attain the Design Eye

TABLE 1

A-4C Airplane Control Reach Survey
 ESCAPAC 1A-1 Ejection Seat
 Distance Beyond Normal Reach to Controls
 (in.)

Controls Requiring Functional Reach	5th Percentile	50th Percentile	95th Percentile
Catapult Handgrip	-	-	-
Manual Fuel Shutoff Lever	-	-	-
Emergency Speed Brake Control	-	-	-
*Air Start Switch	-	-	-
Emergency Landing Gear Release Handle	-	-	-
*Emergency Bomb Release Handle	5.3	1.3	-
Landing Gear Handle (down)	4.8	1.3	-
Emergency Manual Flight Control	3.3	0.6	-
*Emergency Generator Release Handle	4.8	1.3	-
**Canopy Jettison Handle	6.3	1.8	-
Secondary Ejection Handle	-	-	-
Controls Requiring Fingertip Reach	5th Percentile	50th Percentile	95th Percentile
Center Weapons Panel	5.8	-	-

*Critical Controls

**Obstructed by floodlight

NOTE: Total vertical seat adjustment travel 4.25 in. Seat-to-canopy distance; seat DOWN 41.50 in., seat UP 37.5 in. Average differential between normal and maximum possible reach is 1.50 in.

TABLE 2

A-6A Airplane Control Reach Survey (Pilot's Station)
 Martin-Baker GRU-5 Ejection Seat
 Distance Beyond Normal Reach to Controls
 (in.)

Controls Requiring Pilot Functional Reach	5th Percentile Seat Tilted Forward	5th Percentile Seat Tilted Aft	50th Percentile Seat Tilted Forward	50th Percentile Seat Tilted Aft	95th Percentile Seat Tilted Forward	95th Percentile Seat Tilted Aft
Catapult Grip	.8	4.3	-	-	-	-
Brake Selector Handle	-	4.2	-	2.6	-	-
*Electric Ram Air Turbine Handle	-	2.2	-	-	-	-
Spin Recovery Switch	.5	4.3	-	-	-	-
Emergency Flap Switch	-	1.5	-	-	-	-
Manual Canopy Handle	-	4.2	-	1.9	-	-
Hook Release Handle	1.7	6.7	-	4.9	-	1.4
Master Arm Switch	-	4.0	-	1.9	-	-
Landing Gear Handle (down)	1.4	5.8	0.3	4.3	-	-
Controls Requiring Pilot Fingertip Reach	5th Percentile Seat Tilted Forward	5th Percentile Seat Tilted Aft	50th Percentile Seat Tilted Forward	50th Percentile Seat Tilted Aft	95th Percentile Seat Tilted Forward	95th Percentile Seat Tilted Aft
*Emergency Stores Jet- tison Button	1.6	3.1	-	0.7	-	-

*Critical Controls

NOTE: Vertical seat adjustment travel 5.00 in. Seat-to-canopy distance; seat DOWN 45.00 in., seat UP 40.00 in. Average differential between normal and maximum possible reach is 1.70 in.

TABLE 3

A-7E Airplane Control Reach Survey
Douglas ESCAPAC LC-2 Ejection Seat
Distance Beyond Normal Reach to Controls
(in.)

Controls Requiring Functional Reach	5th Percentile	50th Percentile	95th Percentile
* Catapult Grip	2.4	-	-
Emergency Flap Switch	-	-	-
Speed Brake Switch	3.7	-	-
* Emergency Power Handle	7.9	4.5	1.8
* Emergency Generator Switch	5.7	5.8	0.8
Emergency Canopy Jettison Handle	-	-	-
Arresting Hook Handle	- 5.2	- 2.5	-
**Emergency Vent Control	-	-	-
Controls Requiring Fingertip Reach	5th Percentile	50th Percentile	95th Percentile
* Salvo Jettison Button	5.9	2.1	-
***Select Jettison Button	5.1	2.2	-
Auxiliary Jettison Switch	5.2	2.2	-
* Emergency DOWN Lock Release Switch	9.0	5.3	3.0

* Critical Controls

**Obstructed by Canopy Release

***Obstructed by Gooseneck light

NOTE: Vertical seat adjustment travel 5.00 in. Seat-to-canopy distance: seat DOWN 43.5 in., seat UP 38.7 in. Average differential between normal and maximum possible reach is 1.80 in.

TABLE 4

AV-8A Airplane Control Reach Survey
Martin-Baker 9-A MK-1 Ejection Seat
Distance Beyond Normal Reach to Controls
(in.)

Controls Requiring Functional Reach	5th Percentile	50th Percentile	95th Percentile
Emergency Landing Gear Handle	1.5	-	-
*Emergency Trim Switch	-	-	-
Emergency Fuel Jet-tison Switches	-	2.2	-
Controls Requiring Fingertip Reach	5th Percentile	50th Percentile	95th Percentile
*Master Jettison Button	2.2	-	-
Fire Extinguisher Control	-	-	-
Armament Panel	-	-	-
Ram Air Turbine Reset Switch	1.2	-	-

*Critical Controls

NOTE: Vertical seat adjustment travel 4.75 in. Seat-to-canopy distance: seat DOWN 41.50 in., seat UP 37.00 in. Average differential between normal and maximum possible reach is 2.10 in.

TABLE 5

F-4J Airplane Control Reach Survey (Pilot's Station)
 Martin-Baker H-5 Ejection Seat
 Distance Beyond Normal Reach to Controls
 (in.)

Controls Requiring Functional Reach	5th Percentile	50th Percentile	95th Percentile
*Emergency Brake Handle	3.2	-	-
Arresting Hook Control Handle	2.9	-	-
Emergency Vent	3.3	-	-
Canopy Control Handle	-	-	-
Landing Gear Control Handle (down)	4.2	0.7	-
Flap Control Panel	-	-	-
Ram Air Turbine Control Handle	-	-	-
*Catapult Throttle Grip	3.4	0.6	-
Weapons Select Switch	6.7	5.0	5.1
*Secondary Ejection Handle	1.0	-	-
Controls Requiring Fingertip Reach	5th Percentile	50th Percentile	95th Percentile
*External Stores Emergency Release Button	4.1	0.6	-
Missile Jettison Button	3.6	0.2	-

*Critical Controls
 NOTE: Vertical seat adjustment travel 6.00 in. Seat-to-canopy distance; seat DOWN 40.00 in., seat UP 34.50 in. Average differential between normal and maximum possible reach is 2.40 in.

TABLE 6

F-8D Airplane Control Reach Survey
 Martin-Baker MK-P7 Ejection Seat
 Distance Beyond Normal Reach to Controls
 (in.)

Controls Requiring Functional Reach	5th Percentile	50th Percentile	95th Percentile
Emergency Generator Switch	3.2	0.2	-
Arresting Hook Handle	2.9	0.1	-
Emergency Vent Switch	-	-	-
*Emergency Canopy Release	1.9	0.4	-
Emergency Power Handle	2.7	-	-
Emergency Pitch Trim Handle	2.4	-	-
*Emergency Brake Handle	-	-	-
Trim Channel Selector Switch	3.2	0.2	-
Emergency DOWN Lock Release Switch	5.4	2.7	-
Landing Gear Handle (down)	1.3	-	-
Throttle Catapult Handle	0.3	-	-

*Critical Controls
 NOTE: Vertical seat adjustment travel 3.50 in. Seat-to-canopy distance; seat DOWN 40.5 in., seat UP 37.3 in. Average differential between normal and maximum possible reach is 1.50 in.

TABLE 7

OV-10A Airplane Control Reach Survey
 North American LW-3B Ejection Seat
 Distance Beyond Normal Reach to Controls
 (in.)

Controls Requiring Functional Reach	5th Percentile	50th Percentile	95th Percentile
*Emergency Fuel Shutoff Switch	9.0	5.7	3.8
Flap Trim Panel	2.7	-	-
*Stores Emergency Release Handle	9.9	5.8	3.7
Landing Gear Handle (down)	3.6	0.9	-
*Throttles (forward)	2.0	-	-
Controls Requiring Fingertip Reach	5th Percentile	50th Percentile	95th Percentile
Oxygen Regulator Switch	7.1	2.4	0.2
Fire Extinguisher Switch	-	-	-
**Restart Switch	5.2	-	-
Crank Switch	4.9	-	-

*Critical Controls

**Reach obstructed by landing gear handle and throttles.

NOTE: Vertical seat adjustment travel 5.00 in. Seat-to-
 canopy clearance; seat DOWN 46.00 in., seat UP 41.00 in.
 Average differential between normal and maximum possible
 reach is 1.90 in.

Position. This backward and upward travel further removes them from their controls in each airplane surveyed: A-4E, A-6A, A-7E, AV-8A, F-4J, F-8D and OV-10A (Tables 1 through 7).

Recommendations

To achieve maximum effectiveness and safety, aviators should be assigned to present inventory airplanes with the least amount of cockpit/anthropometric incompatibility relative to their own anthropometric dimensions.

Action should be taken to prevent the design and acquisition of future cockpits that will not accommodate the full anthropometric range of potential operators.

In the development of future weapons systems, adequate specifications must be defined in contracts, and these specifications enforced, in order to achieve anthropometric compatibility. During airplane procurement, waivers of detail specifications which involve cockpit design and aviator anthropometry should only be granted by test and evaluation personnel who have the technical expertise to make these decisions.

References

A review of Naval Aircraft Mishaps with human engineering design deficiencies involved. NAVSAFCEN Report of December 1969.

Anthropometry of Naval aviators 1964. NAVAIRENGCEN, ACEL-533.

Anthropometry of flying personnel - 1950. WADC Technical Report 52-321.

A TSD DETERMINATION OF THE TWO-POINT SUPRALIMINAL DLS

ON THE DORSAL FOREARM, THE ANTERIOR THIGH,

AND THE BACK OF THE HAND

Donnell M. Washington

United States Air Force Academy

This investigation was designed to determine the relationship between the tactual two-point threshold and the two-point supraliminal difference limen (DL). The yes/no method was employed throughout the experiment, and three localities were tested. The results confirmed the hypothesis that as the initial two-point threshold increases, the two-point supraliminal DL also increases. No significant differences were obtained between the right and the left sides of each area tested, and only on the hand was a reliable difference obtained between the first nine and second nine sessions. This difference was interpreted in terms of a peripheral fatigue phenomenon.

Although the tactual DL phenomenon has not been explored in any great depth, investigation of this area may provide answers to problems which are encountered in equipment-design research. For example, Fitts (in Stevens, 1951) stated that the use of vision and audition for complex control tasks could possibly reduce the heavy workload invariably carried by the eyes. As the complex equipment utilized in various tasks usually does require considerable input to the sense modalities of vision and audition, it may be desirable in designing future equipment to provide more tactual input to the human operators. In light of this fact, the present investigation of the tactual difference limen (DL) phenomenon may suggest tactual components which could be employed in complex man-equipment systems.

One of the first scientists to investigate the skin as a sense organ was Weber (in Boring, 1942), and one of the issues that Weber addressed himself to was the two-point threshold. Weber discovered by use of a compass that when two points are placed on the skin simultaneously, the points have to be separated a certain distance to be perceived as two points. As the distance is decreased, the probability increases that the points will be perceived as one point. Investigators commonly define the threshold as the distance where two points are perceived fifty percent of the time as "two," and the two-point DL as the distance two points have to be separated in order to be distinguished from another two points applied to the skin.

One problem of determining the subject's sensitivity to a given stimulus is the response set of the subject. Many of the traditional psychophysical experiments are not valid measurements of sensitivity because they fail to adequately measure the subject's attitudinal and motivational variables. In order to alleviate this methodological shortcoming, Green and Swets (1966) urged the application of signal detection theory (TSD) in psychophysical experiments. With TSD the experimenter can measure the subject's response criterion, a catchall term for the attitudinal and motivational variables that affect the subject's decision. The ability to measure a subject's response criterion as well as his sensitivity on a particular dimension is the major advantage of the TSD method.

Recently, Cross, Boyer, and Guyot (1970) obtained supraliminal DLs for four observers. The experimenters employed the yes/no TSD method. The standard stimulus which was always presented first was 47 mm., and the signal stimuli were 50, 53, and 56 mm. After the standard stimulus was presented, it was followed by either the standard or a signal stimulus. The area tested was the dorsal forearm. Each subject (S) received *a priori* information about the distribution of standard-standard and standard-signal trials which was 0.50. The S was instructed to report whether the second stimulus of a trial was the same as or different from the first stimulus of the trial. From the results of the study the investigators concluded that the two-point DL for the dorsal forearm was in excess of 6 mm.

Following the same procedures as Cross *et al.* (1970) Boyer, Cross, Guyot, and Washington (1970) measured DLs using two-point aesthesiometers on the back of four Ss. The standard stimulus was 80 mm. with signals of 85, 90, and 95 mm. All stimuli were supraliminal. Although there were individual differences, all DLs were between 10 and 15 mm. Since the two-point threshold for the forearm and back are 40 and 68 mm. respectively, and as the DLs for these areas are about 6 and 10 - 16 mm., the implications are clear that as the two-point threshold increases so may the two-point supraliminal DL. Accordingly, the aim of this experiment was to determine the relationship between the two-point threshold and the DL.

Method

A 23-year old female served as a subject in this experiment. Four 4-inch aesthesiometers, a pair of opaque goggles, and a vaporizer were utilized throughout the investigation. Furthermore, the yes/no TSD method, a binary detection task, was employed. Each trial consisted of a warning interval, an observation interval, and an answer interval. The warning stimulus, which consisted of a verbal alert of "ready" by the experimenter, prepared the subject for the observation interval.

The standard stimuli were 20% greater in magnitude than the absolute limens reported by Boring (1942) and Hilgard (1962), and the three signal stimuli for each locality were in proportional increments to the particular standard. The standard and signal stimuli for each locality are presented below:

Locality	Absolute Threshold	Standard	Signals		
			(1)	(2)	(3)
Forearm	40.0mm.	48mm.	51mm.	54mm.	57mm.
Anterior Thigh	67.5mm.	81mm.	86mm.	91mm.	96mm.
Back of the Hand	31.5mm.	38mm.	40mm.	42mm.	44mm.

Each trial consisted of always presenting the standard stimulus first and following it with either the standard stimulus again or with a signal stimulus.

The S was instructed to respond either "same" (I have not detected a difference between the two stimuli) or "different" (I have detected a difference between the two stimuli). For each locality there were nine practice sessions and eighteen testing sessions. However, only the data collected during the testing sessions were used in computing the DLs. During all the sessions S wore opaque goggles in order to mask visual cues, and a vaporizer was turned on at least three minutes prior to the beginning of each session to eliminate any subtle auditory cues. The S was given feedback after each trial during the practice sessions to help her select a criterion which would maximize her "hits" and minimize her "false alarms." A description of "hits" and "false alarms" may be found in Green and Swets (1966).

Each session consisted of six blocks of 20 trials each or a total of 120 trials. Two-minute rest intervals were permitted between each 20-trial block except between blocks three and four during which S was given a ten-minute rest interval.

Within each block there were ten standard trials and ten signal trials which were presented at random. The blocks were balanced throughout the experiment, and each signal stimulus was presented in each half of a session. Furthermore, signal previews, which consisted of six trials in which S had knowledge of the results, preceded each block of trials. Three of the trials were standard-standard and three of the trials were standard-signal.

The S was given information about the *a priori* distribution of standard and signal trials. The probability of standard-standard or standard-signal was always equal to 0.50. In addition, because a fatigue effect was found in the results obtained on the hand, three additional sessions were run on the hand one week after the completion of the study.

Results

All data were analyzed using binomial ellipses in Receiver-Operating-Curve (ROC) unit squares. The procedures for analyzing the data are identical to those employed by Cross *et al.*, Boyer *et al.*, and are discussed in Green and Swets (1966). Furthermore, the subject's response criteria (β) for several different experimental conditions were completed by using the methods outlined in Corso (1967).

The upper two panels and the lower left panel of Figure 1 show the difference in sensitivity between the first nine sessions and the second nine sessions on the back of the hand; in addition the panel contains three supplementary sessions taken on the hand. Furthermore, Table 1 presents the sensitivity parameters; Table 2 shows the criteria (β s) employed by S throughout the experiment, and Table 3 summarizes the DLs obtained for each locality.

Although 360 trials were used for both the "false alarm" and "hit" rates of each point in a ROC square, a sample size of 250 was used to compute the 95% of confidence bands. The use of 250 made the statistical test conservative.

In the test of sensitivity of the forearm, thigh, and hand shown in Figure 1, S was reliably sensitive at all three signal levels for each locality. Furthermore, the stimuli were monotonically related for each locality, but the increases in sensitivity from the smallest to the largest signal were not significant.

The test of the first and second nine sessions of the hand revealed a significant difference between the two ellipses. The S was sensitive during the first nine sessions and was not during the second nine. The analysis for these data appear in the lower right panel of Figure 1.

The ROC point of the three additional test sessions were computed from 180 trials for each of the "false alarm" and "hit" rates. However, to make the test conservative, a sample size of 100 was used to compute the 95% confidence bands. Although the ellipse of the additional sessions touched the ellipses of both the first and second nine sessions, the additional ellipse did not touch the chance line. The results of the additional sessions for the hand are shown in the lower right panel of Figure 1.

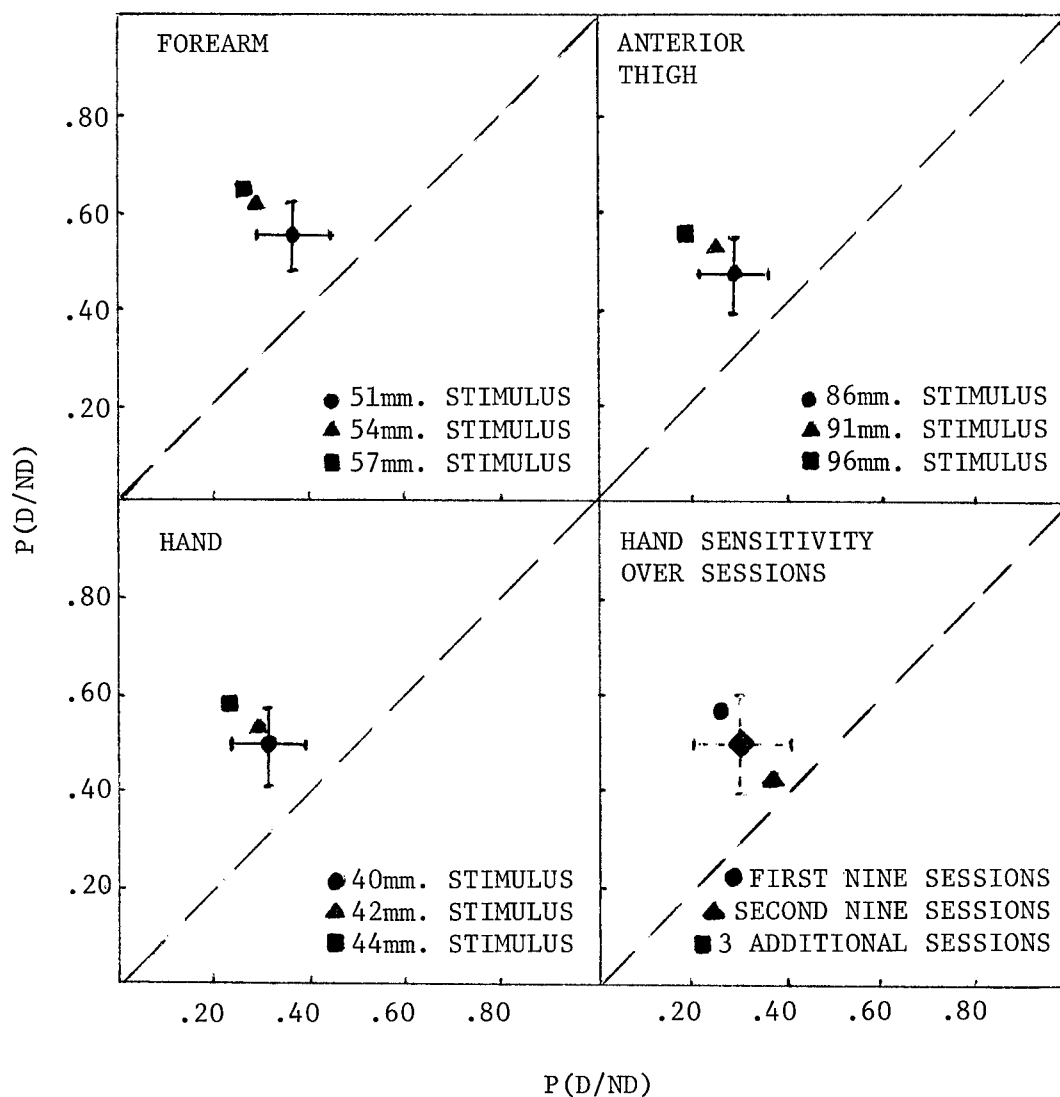


Fig. 1. ROC sensitivity.

TABLE 1
Sensitivity (d') Values for Each Locality

Source	Forearm	Thigh	Hand
Smallest Signal	.23	.28	.25
Middle Signal	.56	.46	.39
Largest Signal	.64	.64	.56
Right Side	.51	.46	.44
Left Side	.43	.49	.38
First Nine Sessions	.38	.54	.69
Second Nine Sessions	.56	.37	.10
Additional Sessions	---	---	.46

TABLE 2
Criteria (β) Values for Each Locality

Source	Forearm	Thigh	Hand
Smallest Signal	1.00	1.03	1.04
Middle Signal	1.02	1.04	1.05
Largest Signal	1.01	1.06	1.08
Right Side	.99	1.04	1.08
Left Side	1.03	1.06	1.04
First Nine Sessions	1.02	1.02	1.07
Second Nine Sessions	1.00	1.07	1.02
Additional Sessions	----	----	1.04

TABLE 3

DL and Absolute Two-Point Threshold for Each Locality

Locality	Absolute Two-Point Threshold	DL
Hand	31 mm.	1-2 mm.
Forearm	40 mm.	3 mm.*
Thigh	68 mm.	4-5 mm.

*As the ellipse for the smallest signal of the forearm just barely missed touching the ROC chance line, the DL was actually slightly less than 3 mm.

Since the binomial ellipses for all three of the smallest signals of each locality were close to the chance line and as their d' values were all relatively low, the DLs were estimated for each locality and are described in Table 3. In addition as summarized in Table 2, S employed a response criterion (β) essentially at an optimal level throughout the experiment. For a description of an optimal response criterion see Green and Swets (1966).

Discussion

The results of the present investigation confirm the hypothesis that as the absolute two-point threshold increases, the two-point supraliminal DL also increases. At each locality, moreover, the stimuli were monotonically related--a relationship predicted by the signal-detection theory. Comparing the sensitivity parameters for the smallest signal of each locality revealed that the d' values were essentially identical even though the signals represented different distances from their respective standards (forearm--3mm., thigh--5mm., and hand--2mm.). This trend was also prevalent for both the middle and the largest signals. Hence, these results indicate that sensitivity to two-point changes is proportionately related to the initial two-point threshold.

The DL of about 3mm. obtained in this study differs from the DL of about 6mm. obtained in the Cross *et al.* experiment on the dorsal forearm. Since one of the four Ss in the Cross *et al.* study was sensitive at the smallest signal (3mm.), the variation in difference limens between the two studies could be attributable to individual differences.

Another possibility is that the lower sensitivity found in the Cross *et al.* experiment was due to peripheral fatigue, especially in light of the fact that the subjects were given 96 trials per session without rest periods. It is also conceivable that the warning stimulus used in the present investigation may have functioned to increase S's sensitivity to the signals. Treisman (1964) employing a yes-no TSD method demonstrated that when an accessory stimulus (light) was regularly presented prior to a critical stimulus (auditory), Ss were more sensitive to the critical stimulus than when the accessory stimulus was irregularly presented. It is possible, therefore, that a warning stimulus could have the same effect on the tactual DL.

Thus any one or all of the forementioned factors could affect the two-point supraliminal difference limen. In any case, further investigation is necessary to determine how (if they do) these factors influence the difference limen.

Boyer, *et al.* (1970) obtained a DL for the back, which has an initial two-point threshold of 68mm., between 10-15mm. These results between the present investigation and the Cross, *et al.* experiment are also applicable in explaining the different findings between this study and the Boyer, *et al.* experiment.

An important finding of the present investigation was that as the magnitude of the signal increased the "false alarm" rate for the signal decreased. This result is predicted by TSD. In the Cross, *et al.* and Boyer, *et al.* experiments.

When ROC tests were performed on the right and left sides of each locality, no significant results were obtained between sides. Moreover, each side of each locality was significant from the ROC chance diagonal. These results suggest that for a specific locality, sensitivity is independent of the side tested.

Furthermore, on the forearm and anterior thigh there were no significant differences obtained between the first nine sessions and the second nine sessions. Hence, there was a lack of evidence to support a practice or fatigue effect for these two locations. The lack of a practice effect is contrary to the results obtained by the Crammer and the Tawney experiments (in Boring, 1942) on the two-point threshold but supportive of the predictions of TSD (Green and Swets, 1966) that sensitivity does not increase with practice. There was, however, a very large statistical difference between the first and second nine sessions in terms of sensitivity on the hand. Further, the binomial ellipse for the second nine sessions touched the ROC chance line.

Since the anterior thigh and the forearm areas tested were 75 and 60 sq. cm. and on the hand the area was 35 sq. cm., there was a greater probability that the aesthesiometer points touched the same

areas more frequently on the hand than the other locations. Another possibility is that since the skin of the hand is not as thick as the skin of the anterior thigh or the forearm, peripheral fatigue on the hand may have occurred much more rapidly.

In an attempt partially to explain the differences noted on the hand S was given three additional sessions after a week's rest from testing. Although the binomial ellipse for the additional sessions touched the ellipses of both the first and second nine sessions, it was closer to the first half than the second. Moreover, it was also significant from the chance line. These results, therefore, are consistent with the peripheral fatigue hypothesis.

One final but relevant finding of this study was that S employed response criteria close to an optimum criterion throughout the experiment. In other words S said "same" and "different" with about equal frequencies; S's strategy was apparently to maximize her "hits" and to minimize her "false alarms." Probably, the *a priori* information, the practice sessions, and the signal previews served to anchor S's criteria at an almost optimal level.

In conclusion, the results of this experiment have supported the hypothesis that as the initial two-point threshold increases, the two-point supraliminal difference limen also increases. Moreover, since the subject consistently employed response criteria almost identical to an optimal level, this study can be used as a basis for further research of the tactual DL phenomenon.

References

- Boring, E. G. *Sensations and Perception in the History of Experimental Psychology*. New York: Appleton-Century-Crofts, Inc., 1942.
- Boyer, W. N., Cross, H. A., Guyot, G. W., & Washington, D. M. A TSD determination of a DL using two-point tactual stimuli applied to the back. *Psychonomic Science*, 1970, 21, 195-196.
- Corso, J. F. *The Experimental Psychology of Sensory Behavior*. New York: Holt, Rinehart, and Winston, Inc., 1967.
- Cross, H. A., Boyer, W. N., & Guyot, G. W. Determination of a DL using two-point tactual stimuli: A signal detection approach. *Psychonomic Science*, 1970, 21, 198-199.
- Green, D. M., & Swets, J. A. *Signal Detection Theory and Psychophysics*. New York: John Wiley & Sons, Inc., 1966.

Hilgard, E. R. *Introduction to Psychology*. New York: Harcourt, Brace & World, Inc., 1962.

Stevens, S. S. *Handbook of Experimental Psychology*. New York: John Wiley & Sons, Inc., 1951.

Treisman, M. The effect of one stimulus on the threshold for another: An application of signal detectability theory. *The British Journal of Statistical Psychology*, 1964, XVII, 15-35.

PERFORMANCE AS A FUNCTION OF TASK DIFFICULTY

IN A CRESPI REVERSAL SITUATION

Lawrence F. Sharp and Gregory D. Smith

United States Air Force Academy

Twenty adult male albino rats were used to validate and extend the Crespi Effect. Subjects were placed in operant chambers and trained to bar press for varying amounts of weight on the bar. Dependent measure was the number of bar presses on an FR schedule. The elation effect was shown in the increased performance after reversals from more to less difficult tasks and the depression effect was evidenced in the opposite direction. The performance of the subjects on the least difficult task was significantly greater than of those with the most difficult task.

In 1942, Leo Crespi published the results of his extensive investigations on the "Quantitative Variation of Incentive and Performance in the White Rat." Crespi was principally concerned with the answers to three questions:

1. What is the relationship between magnitude of incentive and the level of performance?
2. What is the relationship between magnitude of incentive and the gradients within performance?
3. What are the effects of variation of magnitude of incentive upon level of performance?

You may recall that, to answer these questions, Crespi trained groups of white rats to traverse a straight runway to obtain varying amounts of solid food reward. His dependent measure was time to negotiate the runway. The runway was 20 feet long.

In the first of his series of experiments, Crespi attempted to ascertain whether or not there was a difference in performance under conditions of varying amounts of food. His results are shown in Figure 1. Clearly, speed of running was directly related to amount of food available.

In the second of his experiments, Crespi trained all the animals to traverse the runway for a 16-unit reward until running speed stabilized and reached asymptote. He then shifted part of the group to a 64-unit reward and the others to a one- and 4-unit reward,



Fig. 1. Relation of speed of running and amount of food available in Crespi's original work.

while holding others at the 16-unit level. Again, running speeds were ordered according to the amount of incentive, and Crespi observed what he termed "contrast effects." An upward shift in amount of incentive resulted in performance significantly superior to the level of performance of rats receiving the same amount, but who had not had prior adaptation to the smaller amount. He labeled these as "elation effects." Conversely, a downward shift in amount of incentive resulted in "depression effects." That is, their level of performance became significantly *inferior* to the level of performance of those rats receiving the small amount of incentive, but who had not had prior adaptation to the larger amount. Traditional graphs of these effects have shapes as shown in Figure 2. Crespi interpreted these elation and depression effects as experimental evidence for defining a variable within a rat which, in analogous human terms, might be called "expectation." That is, attainment of amounts of incentive below the level of expectation is frustrating; and the attainment of amounts above the level of expectation is elating. Crespi admitted that these terms, i.e., elation and frustration, were based on qualitative observation of rat behavior; but he unequivocally labeled the depression effects as due to "frustration." For both effects he concluded that with drive level held constant, performance is not determined by quantity alone, but also by preceding experiences with quantities.

Following Crespi's work, other researchers including Zeaman (1949), Spence (1954), Metzger, et al., (1957) and O'Connor & Claridge (1958) searched for the contrast effects in wide varieties of tasks in diverse organisms. These results were inconsistent enough to cause a number of researchers, principally Cofer & Appley (1964), to question Crespi's

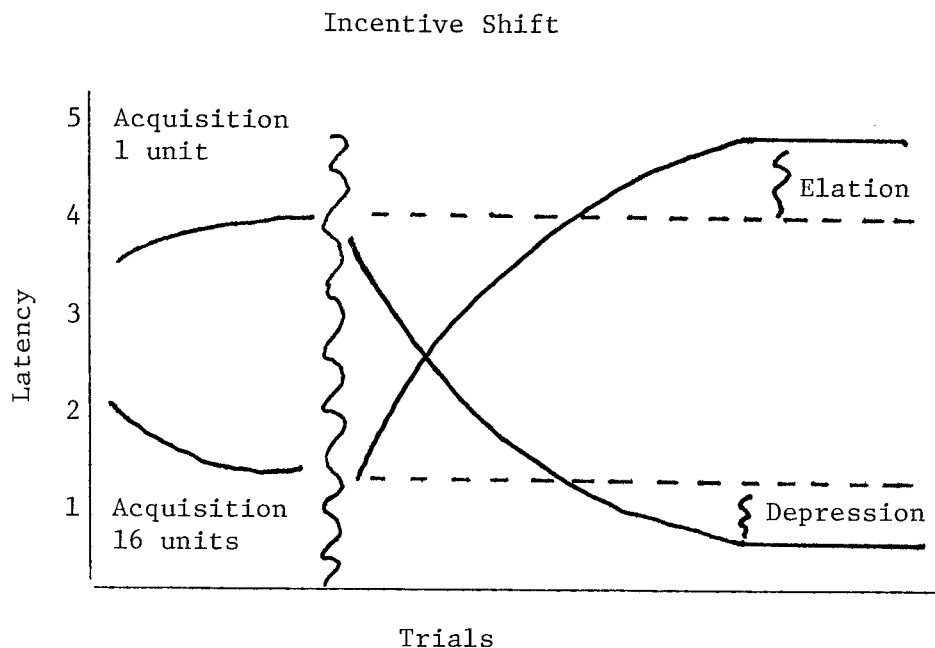


Fig. 2. Crespi's initial basis for denoting elation and depression effects resulted from these findings in 1942. Shown here are the graph's traditional shapes.

conception of the phenomenon.

For a number of reasons we became interested in this phenomenon. First, there was the challenge of replicating a rather significant piece of research in the history of psychology. Second, we wanted to find out if the reversal of performance and contrast effects could be carried beyond one alteration; and last, would the effects show up using different variables?

Method

We made variable substitutions as follows: in place of the independent variable of incentive amounts, we used an FR 2 bar press, with the bar loaded with 1.5, 2.75 and 4 oz. of lead. The dependent variable was, therefore, obviously bar press rate.

Twenty-five, 120-day old male, Sprague-Dawley albino rats were used. There were 20 in the experimental group and 5 were retained as weight controls. After receipt from the supplier, the 20 experimental animals were placed on a food deprivation schedule which lasted two weeks. This was designed to reduce and maintain their body weight at a mean of 85% of what it would have been had they been allowed *ad lib* access. We, of course, determined this by daily comparison with controls and concurrent adjustment of food intake. At the end of the two week period, weights had ranged from 84.6 to 86.3% of the control

mean weight. After deprivation weights were stabilized, all experimental animals were bar press trained on a CRF schedule in BRS operant chambers with no-load bars. Following this, animals were randomly assigned to the 4 experimental groups and bar press activity continued with the groups operating bars loaded with 0, 1.5, 2.75, and 4 oz. of lead, respectively. This training was continued until performance stabilized.

Results

Results of the pre-shift training are shown in Figure 3. As you can see, pre-shift performance is ordered approximately in terms of bar load weight. This is similar to Crespi's running speed wherein his rats were ordered according to incentive amount.

Following this, the first reversal was initiated and each group bar pressed at their new bar load weights for *one* hour per day for three days. At the end of this 3-day period, the second reversal was initiated following the same procedure. Results of the post-shift training for Groups II and IV are shown in Figure 4.

We were principally interested in looking, as Crespi did, at differences arising solely as a function of a shift upward or downward in the independent variable--in this case, bar load weights. Mean group performance during the training (pre-shift) period differed significantly from both the first and second reversal periods, and group performance during the first reversal differed significantly from the second reversal. These are exactly analogous to Crespi's results, and we have carried it an additional alteration.

We also analyzed within group differences for both the first and second reversal. As Figure 4 depicts, in both cases Group II differed significantly from Group IV at the .01 level.

Discussion

A fair question is, "What happened to Groups I and III?" A fair answer is, "We are not quite sure." Figure 5 indicates their pre- and post-shift performance. As you can see, as the Group III bar weight was shifted, their performance was in the direction predicted by the Crespi effect; but it certainly does not show the elation and depression effects as did Groups II and IV. In the case of Group I, as their bar weight was changed from 0 to 4 oz. and back to 0, their performance steadily rose. Remember, Group I rats were those initially trained with a no-load bar. We have no explanation for their steady increase except to cite Crespi's observation that his rats ran significantly faster to no incentive at all than to a very small incentive. Whether we can legitimately equate no incentive to a no-load bar simply because performance increases, is, of course, open to question. The parallel is very interesting though.

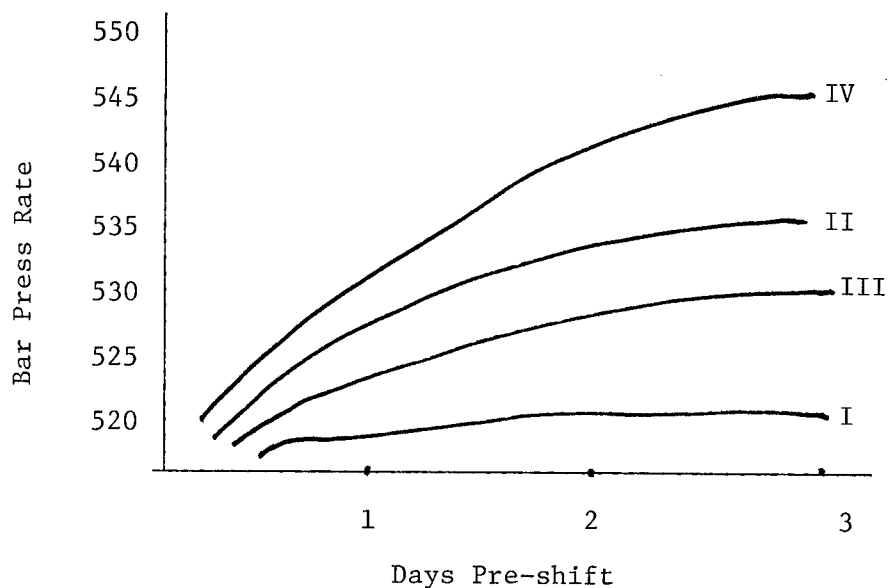


Fig. 3. Results of pre-shift performance training is shown here. The four groups were trained until performance stabilized.

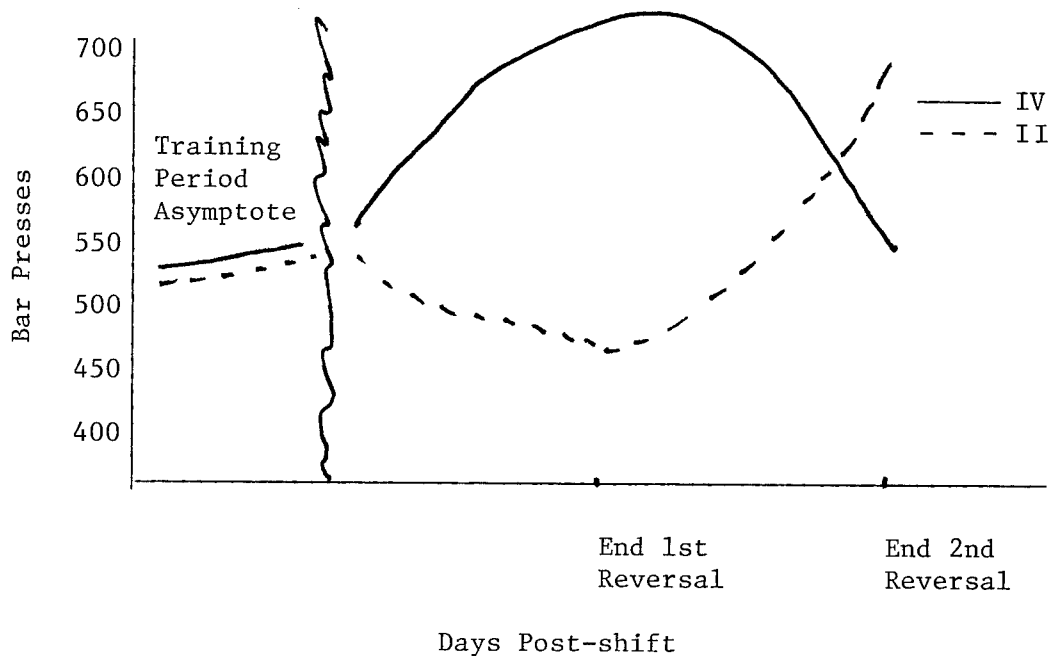


Fig. 4. Results of post-shift training for Groups II and IV.

Nevertheless, we feel we have provided experimental evidence for answers to at least a portion of the questions we set out to investigate;

1. The Crespi Effect can be generalized to a task difficulty variable in addition to incentive amount.
2. A single performance reversal is not a unique behavioral phenomenon, but can be manipulated beyond one alteration.

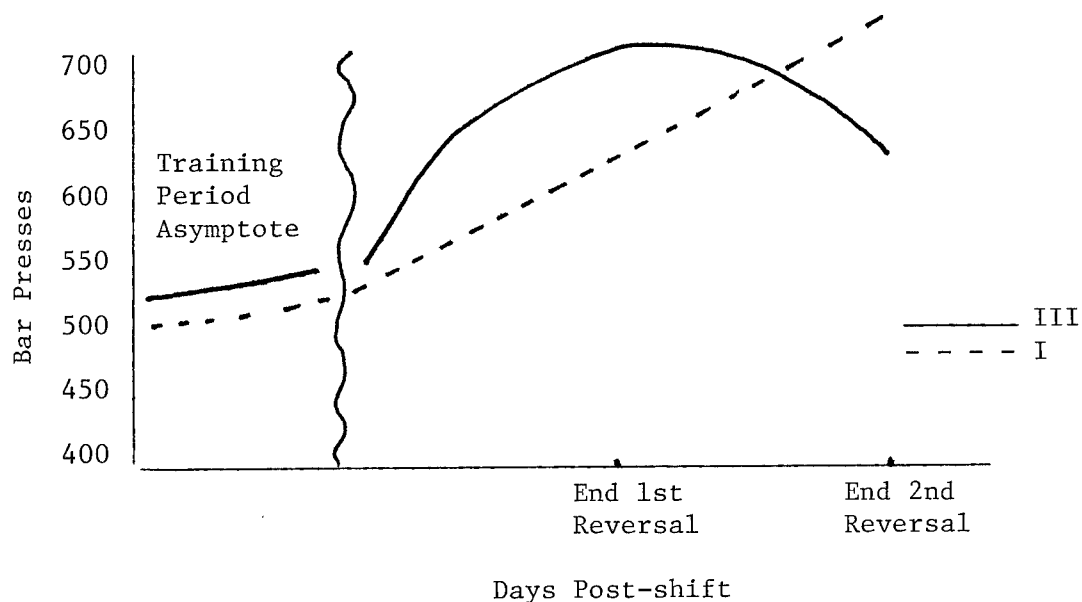


Fig. 5. These are the results of post-shift training for Groups I and III.

TABLE 1
Analysis of Variance

Source	Reversal 1	Reversal 2
Reversal Effect	10.11*	29.66*
Interaction	6.71	2.14
Task Difficulty	7.52*	1.40

$df = 1,112$
 $p < .01$

References

- Cofer, C. N., & Appley, M. H. *Motivation: Theory and research*. New York: John Wiley & Sons, 1965.
- Crespi, L. P. Quantitative variation of incentive and performance in the white rat. *American Journal of Psychology*, 1942. 55, 467-517.
- Metzger, R., Cotton, J. W., & Lewis, D. J. Effect of reinforcement magnitude and order of presentation of different magnitudes on runway behavior. *Journal of Comparative and Physiological Psychology*, 1957, 50, 184-188.
- O'Connor, N., & Claridge, G. S. A "Crespi Effect" in mole imbeciles. *British Journal of Psychology*, 1958, 49, 42-48.
- Spence, K. W. *Behavior theory and conditioning*. New Haven: Yale University Press, 1956.
- Zeaman, D. Response latency as a function of the amount of reinforcement. *Journal of Experimental Psychology*, 1949. 39, 466-483.

PRIVATE MOTOR VEHICLE ACCIDENT RATES OF

USAF PILOTS VS NONPILOT OFFICERS

Anchard F. Zeller and James C. Marsh

Air Force Inspection and Safety Center
Norton Air Force Base, California

An evaluation of the private automobile accident experience of Air Force rated and nonrated officers for a three-year period indicates that rated officers have a consistently lower automobile accident rate than nonrated officers. It is suggested that this difference is the result of the greater mechanical orientation of officers who attain aeronautical ratings. The rigid selection process which rated officers must undergo, together with the frequent physical examinations which they are required to take, might also be a variable although there is no direct evidence to support this possibility.

Although United States Air Force aircraft accidents receive greater publicity and, in terms of dollars, are more expensive, greater personnel losses accrue each year from the much less spectacular ground accidents. Among ground accidents, the most prevalent are those associated with the operation of an individual's own private motor vehicle (PMV). It is quite desirable, therefore, that the variables associated with PMV accidents be evaluated to isolate those which may be of possible accident prevention import.

It has been reported that PMV accident rates decrease in conjunction with increasing age. A concomitant variable is the decrease in accident rate with increasing rank. There appear to have been no studies, however, which relate USAF occupational specialty to the propensity for PMV accidents. Within the USAF structure, the officer cadre is made up of pilots and nonpilots, with almost one-third being in the pilot category. On an *a priori* basis, it would seem that individuals who have both chosen and been selected to operate aircraft would have greater interest in and appreciation for mechanical equipment and, because of continuing exposure to high speed/distance/rate of closure judgments, would be less involved even in PMV accidents than officers whose primary service duty did not include these kinds of interest and experience. It would also appear reasonable that pilots thoroughly indoctrinated in the use of personal equipment and in the necessity for maintaining good personal physical condition would be more inclined to make use of such aids as seat belts. Whether or not the difference between their use of alcohol in conjunction with driving would be less because of the discipline learned in associating with flying is less easy to anticipate. While

their experience would lead to an anticipated lesser alcohol involvement, anecdotal material would suggest that there might be no difference, or even a greater degree of such involvement among pilots involved in PMV accidents. The present study was designed to explore these various considerations.

Method

A large number of variables were encoded and placed on computer tapes for every accident involving USAF personnel. Information relating to all private motor vehicle accidents experienced between 1 January 1968 and 31 December 1971 was retrieved from this source for all officer personnel. Pilots and nonpilots were considered separately. Accident rates for the two groups, computed on the basis of 100,000 people, were developed using USAF personnel strength figures for each of the four years considered.

Results and Discussion

During the period studied, pilots had been involved in 78 PMV accidents, while nonpilots had been involved in 409. When these numbers are converted into rates on the basis of personnel strength, the pilot annual rate of 49 compares with the nonpilot rate of 113. It is apparent that the relative involvement of pilots in PMV accidents is less than that of their nonpilot fellow officers. The results given are the totals for the entire 4-year period. When each of the four years is considered separately, the numbers and rates vary somewhat, but for none of the years considered is there a reversal of the trend of the summary findings.

Although not included in the current evaluation, data from 1959 and 1960 were subjected to the same kind of analysis. The results of this were directly comparable. It appears that there is every reason to believe that USAF pilots consistently experience relatively fewer PMV accidents than do nonpilots.

As a further refinement, the accidents for the four years for both groups were spread by the rank of the individual involved. Among pilots there was a consistent decrease in rate, with second lieutenants having a rate of 99 and generals and colonels collectively having a rate of 6. Each ascending rank experienced a lower rate than the preceding. The same general trend held for nonpilots. Second lieutenants had a rate of 156 and lieutenant colonels a rate of 76. To this point, each ascending rank had a decreased rate; however, the general/colonel category had a rate of 155, only one point less than that of the second lieutenants. While the numbers are small, only 16 in the general/colonel category, it is interesting to speculate that the individuals on flying status in the older age groups who are required to maintain higher physical standards and who continue to

operate in a mechanical world may be better equipped to avoid PMV accidents.

It also appears that the severity of accidents is less for pilots than for nonpilots. Of the 78 pilot accidents, eight involved at least one fatality, for a rate of 5. Of the 409 nonpilot accidents, 76 involved a fatality, for a rate of 21. A Chi-square test indicated this difference to be at the 8 percent confidence level. The trend is substantiated by an analysis of the individual years. In each, pilot accidents involving fatalities were relatively fewer than those of nonpilots. The reason for this is not readily apparent; there are a number of possibilities. One is the possible greater use of such protective equipment as seat belts, and another is the possibility that pilots, again by virtue of their mechanical interests, drive vehicles which are better able to withstand crash impacts.

Private motor vehicles can be classed as either 4-wheel or 2-wheel. Of the 78 pilot-involved accidents, 27 involved a 2-wheel vehicle. Of these, only one was fatal, in contrast to 13 of the 76 nonpilot accidents which involved 2-wheel vehicles. Although the numbers are small, the tendency for pilot-involved accidents to be less severe (as measured by fatalities) is even more pronounced in 2-wheel than in 4-wheel vehicles. Why is not readily apparent. The one pilot fatal 2-wheel vehicle accident involved running off the road. Four of the nonpilot fatal accidents were attributed to this cause. It is interesting to note that eight additional nonpilot fatal accidents involved collisions between motor vehicles. No 2-wheel pilot fatal accidents were in this category.

When all of the PMV accidents, both 4-wheel and 2-wheel, are considered in relation to cause, it is noted that approximately 36 percent of those involving pilots were between motor vehicles, while 47 percent of those involving nonpilots were between motor vehicles. Again, the greater involvement of nonpilots in accidents which require distance/rate of closure judgments tends to support the general hypothesis that pilots whose primary duty involves such judgments are more proficient in avoiding accidents in this category. A slightly larger proportion of the nonpilot accidents also involved collisions with fixed objects. By contrast, noncollisions were slightly more prevalent among the pilots.

Both civilian and military statistics routinely indicate alcohol as a major factor in PMV accidents, with over half of all PMV fatal accidents being associated with alcoholic intake and driving. As would be expected, alcoholic involvement in accidents experienced by officers is consistently less than that in accidents for either the military or the civilian population at large. Twenty-seven of the 78 pilot-involved accidents were associated with alcohol, for a rate of 17. Three of the eight pilot fatal accidents involved alcohol.

Ninety-eight of the 409 nonpilot accidents were associated with alcohol, for a rate of 27. Thirty-one of the 76 nonpilot fatal accidents involved alcohol. While the rate per 100,000 officers is higher among the nonpilots, relatively more of the accidents which pilots do experience are associated with alcohol--approximately 35 percent compared to 24 percent for nonpilots.

Among these alcohol-involved accidents the severity of pilot accidents as measured by fatalities is less. Only three of the 27 (11 percent) involved fatality, in contrast to 31 of the 98 (32 percent) nonpilot alcohol-involved accidents. As in the other evaluations, the surprising consistency of the four years considered individually with the summary findings suggests that this is a valid observation. It is of parenthetical interest to note that direct alcohol involvement in USAF aircraft accidents is almost nonexistent.

It was considered that the severity of nonpilot accidents might be related to the more effective use of seat belts and helmets by pilots who, by the nature of their flying assignments, routinely utilize these kinds of equipment. Although data on all accidents were not recorded, those data which were available indicate that in 4-wheel vehicles pilots used available seat belts 29 times and did not use them 19 times. Nonpilots, on the other hand, used them in 157 cases and did not use them in 142. The relatively greater use by pilots is in the expected direction. Only one pilot fatality in a 4-wheel vehicle involved the non-use of seat belts. Thirty-eight non-pilot fatalities were associated with such non-use.

It is of some interest to note that while in both pilot and non-pilot accidents seat belts were used oftener than not, in alcohol-involved accidents this was not the case. From the information available, in 12 instances pilots did not use seat belts when alcohol was involved in a 4-wheel vehicle accident; they did use them in 11 cases. Nonpilots did not use seat belts in 54 cases of alcohol-involved 4-wheel vehicle accidents; they did use them in 30 cases.

It appears that alcohol makes all individuals, both pilots and nonpilots, less concerned with the use of available safety equipment. Although the information regarding the use of helmets in 2-wheel vehicles is too incomplete to provide critical analysis, the same tendency for a relative decrease in the use of helmets in the alcohol-involved accidents in contrast to those in which alcohol was not involved is apparent.

Conclusions

An evaluation of the PMV accidents experienced by USAF officers, both pilots and nonpilots, for a 4-year period substantiates the hypothesis that pilots are relatively less involved in motor vehicle

accidents than are nonpilots. This trend is consistently demonstrated in each of the individual years as well as in the summary evaluation. Those PMV accidents which pilots do experience are consistently less severe in terms of fatalities than those of nonpilots. The reason for this is not completely clear, although there is some suggestion that it is related to the greater use of protective equipment by pilots. Other possibilities are better physical condition and more crashworthy PMVs. Alcohol is consistently less a factor in all officer PMV fatal accidents than in accidents experienced by the general civilian or USAF population. Pilots have a lower rate of alcoholic involvement than nonpilots, but relatively more pilot accidents than nonpilot accidents have alcohol as an associated factor.

While factual evidence upon which to base a reason for these differences is not readily available, the hypothesis that pilots by the nature of their interests, training, and experience are better able to handle all types of mechanical vehicles appears to be in keeping with the factual findings.

Training (Dual Session)

CHAIRMAN: Major Joseph K. Jarboe

CONFIDENCE TESTS AS DIAGNOSTIC AIDS IN TECHNICAL TRAINING

Wayne S. Sellman

Air Force Human Resources Laboratory

Gary J. Echternacht,
Robert F. Boldt

and

Joseph D. Young

3345th Technical School

Educational Testing Service

One of the primary tasks facing technical training instructors is that of accurately assessing student knowledge of course materials. This study attempted to evaluate the ability of confidence testing, used in conjunction with daily "diagnostic" quizzes, to improve the average level of achievement in courses in technical training. It was concluded that confidence testing could result in improvement in achievement in some segments of courses.

One of the most popular methods of testing student achievement is through the use of multiple-choice test items where the examinee is presented a question and a number of alternatives from which he is to choose the correct answer. However, the notion of requiring an examinee to choose only one alternative from a fixed number has been subject to criticism. For many times the examinee is quite sure as to the correct choice and has no difficulty indicating it; on the other hand, he may be able to eliminate some of the alternatives and then be forced to guess between the remainder. Knowledge is not an all-or-none proposition. It seems reasonable to assume that a student who can eliminate some alternatives has more knowledge than one who can eliminate none, and a student who selects an answer and indicates his doubt as to its correctness has more knowledge than one who is completely misinformed and yet certain of his answer.

One possible approach for providing diagnostic information to instructors is confidence testing (de Finette, 1965; Ebel, 1965; and Shuford, Albert, & Massengill, 1966). Advocates of confidence testing believe that their procedures provide more information and yield "fairer" scores than conventional multiple-choice testing since measures of the level of student knowledge of each test item are acquired rather than a simple indication that the student was right or wrong. Instructors could thus identify the level of student knowledge and consequently more accurately ascertain how and what additional teaching should occur.

If, in fact, confidence testing does provide information concerning a student's level of knowledge beyond that provided by conventional multiple-choice tests, it would appear that its use would allow instructors to tailor course presentations to correct student weaknesses and make materials more meaningful to students thus enhancing classroom experience. The purpose of this study therefore was to examine the ability of confidence testing, as used with daily, diagnostic quizzes, to increase the average level of achievement in courses of instruction.

Procedure

Subjects

Two courses, Aerospace Ground Equipment Repairman (AGE) and Jet Engine Mechanic (JEM), taught by the 3345th Technical School, Chanute Air Force Base, Illinois, were chosen for investigation. Upon course entry, 434 students, 180 in AGE and 254 in JEM, were randomly assigned to a six-hour instructional shift. The AGE course was divided into four nonoverlapping shifts while the JEM course utilized only two shifts. These shifts were designated "A, B, C, and D" in AGE and "A and B" in JEM. Students proceeded through their courses by completing a series of instructional blocks which covered unified areas within the courses and were of either a one or two weeks duration.

Since the experimenters were primarily interested in confidence testing as applied to a multiple-choice format, the daily quizzes used in each course were examined to determine a period where most quizzes were multiple-choice in nature. Blocks 2 and 3 were selected for further study from JEM while Blocks 6, 7, and 8 were selected from AGE.

Method

The effects of three different methods of daily quiz testing on course performance, as measured by end of block examination scores, were studied. Three methods of testing were considered, two experimental confidence procedures and a control procedure. The control procedure consisted of traditional multiple-choice testing with four alternative response items.

One confidence testing procedure, termed "Pick-One", required the examinee to choose the alternative he believed to be correct, exactly as he would in a conventional multiple-choice test, and then indicate on a five-point scale his sureness of his response (Boldt, 1971).

A second type of confidence testing, termed "Distribute 100 Points", approximated the method devised by Shuford & Massengill

(1968). Using this method, the examinee was first required to choose an alternative and record that as being his selected answer. He then indicated his subjective probability of each alternative's being correct by distributing 100 points over the various alternatives.

A student was assigned to a remedial session of two hours following his scheduled class if he performed unsatisfactorily on the daily quiz (usually scoring below 70 percent), had poor performance in the previous block, or showed weakness in practical performance. Instructors received one of two types of directions for use in the remedial sessions. One was to use remediation procedure of the technical school. The other urged special remediation based on the notion that students responding incorrectly with high confidence should receive a different type of instruction than students responding incorrectly with low confidence. Students who were misinformed (wrong answer with high confidence) would go through a two-stage remedial process, first being instructed why their responses were wrong and then why the answer was correct. Students who were simply not informed (wrong answer with low confidence) would go through only a single stage remedial process, being instructed why the correct answer was correct. In this manner, an initial step could be taken to allow instructors to tailor their remedial instruction to the needs of the students.

Results

A three-way factorial analysis was used where independent variables were type of testing, type of remediation, and shift. The dependent variables used in this analysis were the respective end of block examination scores. Since these dependent variables were correlated, a multivariate analysis of variance was used (Morrison, 1967; Pruzek, 1971).

In the AGE course three end of block examination scores served as criteria. When the shifts by testing type interaction was tested for significance, two of the three discriminant functions available were found significant with probabilities less than .05. These results appear in Table 1. This was interpreted to mean that the testing type effect depended upon the particular shift, block, and type of testing that was being examined; hence, no overall main effects were examined for AGE. In order to better understand this interaction, univariate one-way analyses (Winer, 1971) were performed within each shift, with the testing type effects being calculated in each case. From this point on, the discussion of the analysis of end block examination scores in AGE will be presented shift by shift.

The analysis for Shift A indicated that there were no significant differences between the types of testing.

TABLE 1
Multivariate Tests of Interactions using
Wilks' Lambda Criterion

Test of Roots	F	Degrees of Freedom for Hypothesis	Degrees of Freedom For Error	p less than
AGE				
1 Through 3	2.584	18	484.146	0.001
2 through 3	1.891	10	474.957	0.044
3 through 3	0.062	4	456.715	0.993
JEM				
1 through 2	2.585	4	482.000	0.036
2 through 2	0.110	1	241.500	0.740

In the analysis of testing type within Shift B, one significant discriminant function was found. An examination of the univariate F-ratios (Table 2) indicated that a significant difference occurred only in Block 7. The effects, in terms of deviation of means, of the various types of testing are given in Table 3. The group using multiple-choice testing had the lowest average block examination score, while Distribute 100 Points had the highest. No significant differences between the types of testing were found for the remaining blocks.

When Shift C was analyzed, one discriminant function was found to be significant. Univariate analyses on the Block 6, 7, and 8 scores produced significant F-ratios on only the Block 6 scores. Table 3 indicates that multiple-choice testing was definitely the least effective method in this block, while Distribute 100 Points was the most effective method. Since no significant differences were found in Blocks 7 and 8, the testing types were concluded to be equally effective in these blocks.

No significant differences between testing types were found in Shift D.

In JEM the end of block examinations received while in instructional Blocks 2 and 3 served as criteria. There were three types of testing as before, two types of remediation, and two shifts, designated A and B.

TABLE 2
Univariate F Tests

		F	Mean Square	p Less Than
AGE - Shift D	Block 6	.832	37.318	.437
	Block 7	.934	75.493	.395
	Block 8	.767	45.048	.466
AGE - Shift C	Block 6	4.595	206.099	.011
	Block 7	.732	59.178	.482
	Block 8	.736	43.229	.481
AGE - Shift B	Block 6	.581	26.062	.560
	Block 7	4.307	348.293	.015
	Block 8	.182	10.671	.834
AGE - Shift A	Block 6	.744	33.361	.477
	Block 7	.686	55.448	.505
	Block 8	.442	25.992	.643
JEM - Shift B Special	Block 2	3.496	165.707	.032
	Block 3	11.924	443.185	.001
JEM - Shift A Special	Block 2	12.001	568.557	.001
	Block 3	4.686	174.174	.010
JEM - Shift B Control	Block 2	1.092	51.712	.337
	Block 3	2.015	74.899	.136
JEM - Shift A Control	Block 2	5.450	258.176	.005
	Block 3	5.340	198.477	.005

When testing type by remediation type by shift interaction was tested, (Table 1) one discriminant function was found to be significant. In order to better understand this interaction, the analysis was divided so that the types of testing could be examined within the four combinations of shift and remediation type. From this point on, the analysis will be discussed by these four groups.

One significant discriminant function was found when the types of testing were considered in Shift A for classes using special remediation. An examination of the univariate F-ratios yielded significant differences in both blocks (Table 2).

TABLE 3
Effects of Significant Classes

Class		Multiple-Choice	Pick-One	Distribute 100 Points
A	Block 7, Shift B	-6.667	1.958	4.708
G				
E	Block 6, Shift C	-5.354	.521	4.833
	Block 2, Shift A1	-5.618	6.942	-1.324
J	Block 2, Shift A2	-4.028	2.741	1.287
E	Block 3, Shift A2	-2.225	-1.032	3.257
M	Block 2, Shift B1	-4.564	2.191	2.373
	Block 3, Shift B1	-5.561	5.466	.095

The effects of the various types of testing are also given in Table 3. It was apparent that multiple-choice testing was again low, while Pick-One testing was highest.

Two significant discriminant functions were found when the types of testing were examined within Shift A when control remediation was used. Also, the univariate F-ratios were significant for each instructional block. Thus, it was concluded that there were significant testing type effects in each block. In Block 2 multiple-choice testing was again low. In Block 3, however, both multiple-choice and Pick-One were low while Distribute 100 Points was approximately one standard deviation higher.

When the types of testing were analyzed in Shift B for classes using special remediation, two discriminant functions were found to be significant. When the univariate F-ratios were examined, significant Fs were found in both blocks. In each case the mean block grade was lowest for the group using multiple-choice testing. There seemed little to choose between the Pick-One and Distribute 100 Points method in Block 2, while the Pick-One method appeared to be superior to the Distribute 100 Points method in Block 3.

No significant testing type differences were found in Shift B when the control remediation type was used. Thus, it was concluded that there was no difference in the block scores for the groups using the three types of testing in Shift B when the control remediation was used.

Discussion

Two features of confidence testing stand out as a result of this study. First, confidence testing does not necessarily result in greater achievement as measured by end of block examination scores in technical training courses. In AGE only two analyses with significant differences in mean achievement were found among the three types of testing of the twelve analyses (four shifts and three blocks). In JEM the record was a little better; here significance was found in 5 of the 8 analyses. Yet, the fact remains that significance was found in only 17 percent of the analyses for AGE and in 63 percent of the analyses for JEM. The second conclusion standing out was that when significance was found, multiple-choice testing was found least effective. The picture was somewhat clouded with respect to distinguishing between the two types of confidence testing. In AGE, Distribute 100 Points confidence testing resulted in the highest mean end of block examination score in both cases where significance was found. In JEM there was some question whether Pick-One or Distribute 100 Points was superior, as that seemed to depend upon the particular shift and type of remediation. The analyses did seem to slightly favor Pick-One confidence testing.

References

- Boldt, R. F. *A simple confidence testing format*. AFHRL-TR-71-31. Lowry AFB, Colorado: Technical Training Division, Air Force Human Resources Laboratory, July 1971.
- de Finetti, B. Methods for discriminating levels of partial knowledge concerning a test item. *British Journal of Mathematical and Statistical Psychology*, 1965, 13, 87-123.
- Ebel, R. L. Confidence weighting and test reliability. *Journal of Educational Measurement*, 1965, 2, 49-57.
- Echternacht, G. J., Sellman, W. S., Boldt, R. F., & Young, J. D. *An evaluation of the feasibility of confidence testing as a diagnostic aid in technical training*. AFHRL-TR-71-33. Lowry AFB, Colorado: Technical Training Division, Air Force Human Resources Laboratory, 1971. (Also Research Bulletin RB-71-51. Princeton, N. J.: Educational Testing Service, 1971)
- Morrison, D. F. *Multivariate statistical methods*. New York: McGraw-Hill, 1967.
- Pruzek, R. M. Methods and problems in the analysis of multivariate data. *Review of Educational Research*, 1971, 41, 163-190.
- Shuford, E. H., Albert, A., & Massengill, H. E. Admissible probability measurement procedures. *Psychometrika*, 1966, 31, 125-145.
- Shuford, E. H., & Massengill, H. E. *Confidence testing at the officer training school*. Lackland Air Force Base, September 1968. Lexington, Massachusetts: Shuford-Massengill Corporation, 1969.
- Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1971.

PRELIMINARY STUDIES OF A BETTING MODEL

IN ACHIEVEMENT TESTING TO PREDICT

TROUBLESHOOTING SUCCESS IN THE LAB

Anna May Kelleher, Gordon A. Parrish, and Douglas Adair

United States Coast Guard Training Center

A betting model for indicating answers was tested with two classes of Electrician Mate trainees to predict troubleshooting ability. Trainees were given points to spread or bet on four-response multiple-choice questions on basic electricity. The trainees who bet most frequently on one choice were found to be more proficient in the school troubleshooting examinations.

Success in problem solving can be seen as a personality trait different from IQ and alterable by practice. Karlin, et al (1967) defined creativity as the ability to generate questions and responses and then studied it in a problem solving context. He concluded that IQ was not related to creativity. In a learning experiment conducted by Eiferman (1965) the majority of Ss became systematic in their response patterns after being exposed to a few problems of the same type. Problem solving as a personality trait has been studied by Wikoff (1966), who found response style significantly related to personality differences. Differing personality types have been studied in gambling situations by Edwards (1966). He defined two main groups; one, those who take or avoid big risks and two, those who had preferences for certain probabilities. These differences were noted in real situations vs test situations. Gentile & Schipper (1966) found that by manipulating the odds or difficulty of learning two independent events and a subsequent decision making task they could influence the performance of different Ss. If these gambling categories of Edwards' can be identified in a population then perhaps solving success can be identified by relating it to these gambling styles, i.e., do certain types of gamblers do better in problem solving situations?

Coast Guard Problem

Students were selected by Naval Battery Tests which could not include numerous personality variables, even if they could be identified. Within a school, certain personality traits, such as problem solving styles may contribute much to success or capability of a man in school and/or in the field. Troubleshooting electronic equipment is a practical form of problem solving which is critical to field performance. Instructors often remark that a student who does well on written examinations often is all thumbs, if not suicidal, on handling equipment. These students not only lack mechanical ability

but seem unable to exhibit the prowess they did on a written test when faced with an actual situation. They cannot apply the theory.

Troubleshooting may be something other than IQ or mechanical skill and perhaps reside more in a constant personality trait related to decision making. The decision making situation of troubleshooting may be easier for a person who prefers to take a chance, i.e., to bet all on one answer; and harder for another person who may be reluctant to bet and prefer to spread his bets across the possibilities. This betting pattern may operate independently of knowledge of electronics theory. It was within this framework of risk taking that these preliminary studies were undertaken.

It was hypothesized that: (a) given an opportunity to bet on a multiple choice examination the students would divide into different types of gamblers, (b) those gamblers who bet all on one choice consistently would do better on the practical examinations given on troubleshooting, and (c) these betting patterns would be better predictions of success in troubleshooting than knowledge scores on electrical theory examinations.

Method

Two classes of Electrician Mates (EM) students at Coast Guard Training Center, Governors Island, New York were the subjects. Class 1-11 consisted of 21 students; Class 15-11 of 22. Their Naval Battery Tests entrance scores were homogenous and average age was 19.6 years and average education was 12.05 years. A thirty-question multiple-choice test on basic electricity was administered to both classes two weeks apart. The material had been covered in a Programmed Instruction utilized for homework in Week One of EM school. The administrator of the test was a Petty Officer Third Class of Systems Section with a BA in psychology. He said the purpose of the test was to evaluate a new testing method. The students were instructed that they had 100 points to bet on each question. They could bet it all on one choice or spread it out. A guess meant 25 bet on each choice. Total score would be the sum of amount on each right choice. The highest score was 3000 if 100 was always bet on the right answer. The students were informed that results would be given to the school chief.

The number of single bets of 100 were tabulated for each and also the frequency of 2, 3, and 4 choices.

	type of choice	frequency
Examples:	1	22
	2	5
	3	1
	5	2

Students were ranked by number of one choice bets and also nearness to one choice.

Examples: #8

 #9

type of choice	frequency	type of choice	frequency
1	22	1	22
2	6	2	5
3	2	3	0
4	0	4	3

Student #8 would have a higher rank than student #9 because 6 vs 5 times he split his bet on two choices. An attempt was made but a method was not devised to evaluate a criterion for size of bet being split.

Students were also ranked by scores on troubleshooting exams given in Weeks 11 and 12 of EM school. Comparisons utilizing rank order coefficients were computed between (a) rank on Betting Pattern vs rank on Practical Troubleshooting Test, and (b) rank on Electrical Theory Test vs rank on Practical Troubleshooting Test. Because knowledge or familiarity with material would influence the frequency of single choices, familiarity with material was investigated by testing Class 1-11 in Week 6 of the EM curriculum and Class 15-11 in Week 1.

Results

Table 1 presents the Rank Correlation Coefficients obtained. For Class 1-11 (which took the examination in Week 6 of EM school) the betting pattern rank was significantly correlated with practical exam scores indicating a relationship which was better than the nonsignificant results of theory to practical. For Class 15-11 the theory vs practical coefficient was significant, while the betting vs practical was not.

A graph composed of the cumulative frequency of single choices showed a sharper rise for Class 1-11 than for 15-11, i.e., a difference in betting styles, (see Figure 1) with an interquartile range of 3 for Class 1-11 and 6 for Class 15-11.

Discussion

The significant correlation of .54 for Class 1-11 indicates that Betting Pattern of odds preference is a better predictor of troubleshooting ability than the electrical theory test. The failure to achieve similar correlation for Class 15-11 indicates that knowledge

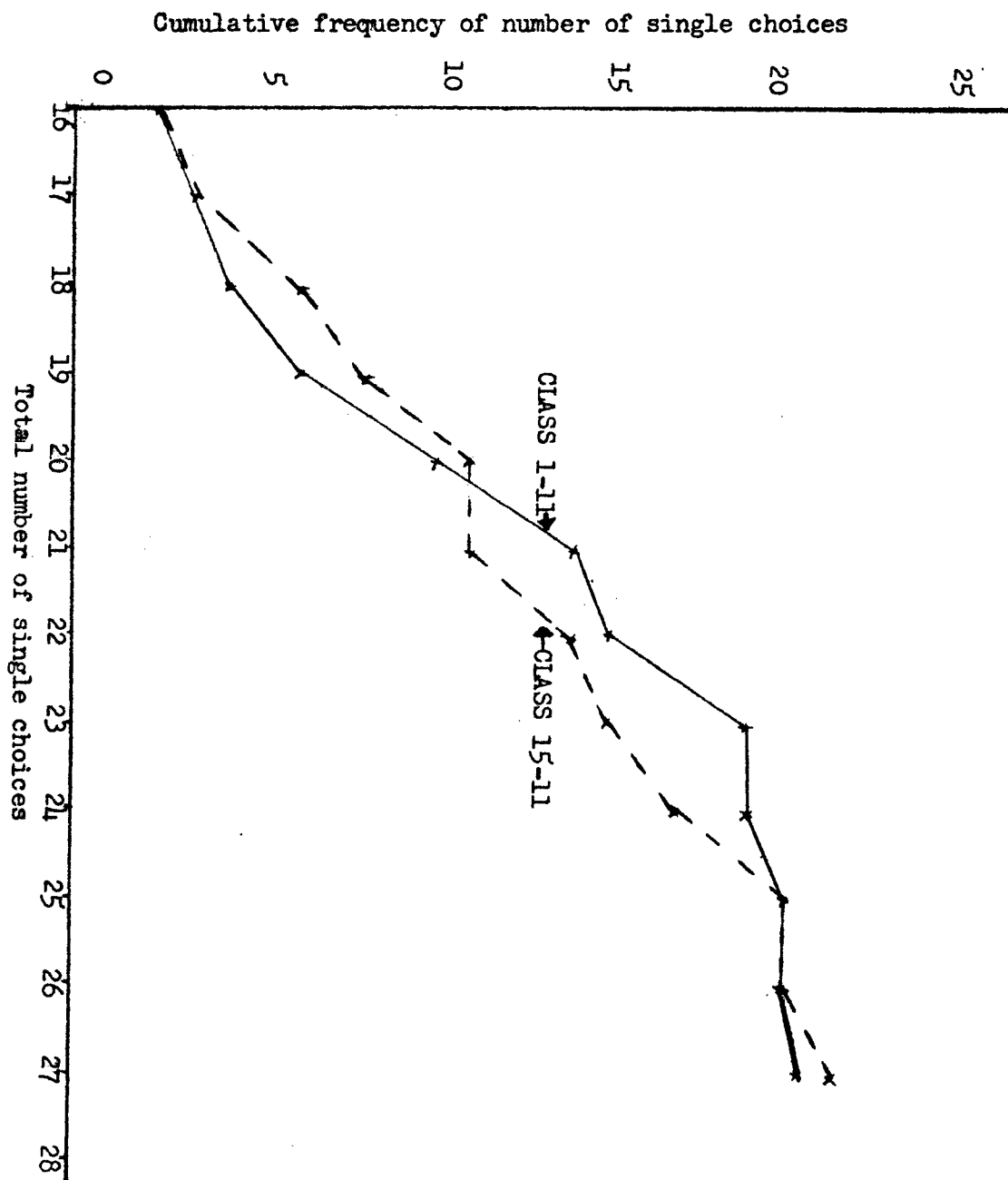


Fig. 1. Cumulative betting pattern curves for both classes.

TABLE 1
Rank Correlation Coefficients

	Class 1-11 n = 21	Class 15-11 n = 22
	r	r
Rank on Betting Pattern vs rank on Troubleshooting	.34	.89*
Rank on Electrical Theory vs rank on Practical Troubleshooting Test	.54*	.21
* $p < .01$		

of theory operates differently in this case. It is theorized that since Class 15-11 took the examination based on material studied in the same week it was more apt to choose the one correct answer; while Class 1-11 which took the examination five weeks later was guessing at material and exhibited a betting style more. It is this class, 1-11, which showed a separation of betting styles within the class, while Class 15-11 showed no such separation. When this betting style was separated from theory knowledge, as in Class 1-11, it was a better predictor of success on the practical application of electrical theory than a score of answers correct.

If the betting model is to be utilized, a neutral test which measures or maximizes betting style must be developed. At present, an examination of Coast Guard Regulations is being tested for its use in spotting betting styles.

In conclusion, therefore, for Class 1-11: (a) a betting model did separate the class by gambling types, (b) this betting model was a better predictor of troubleshooting success, and (c) those Ss who bet most frequently on one choice were better troubleshooters.

Such information has potential in counseling students early in a rating emphasizing troubleshooting and designing learning experiences which maximize gambling styles.

References

- Edwards, P. Probability preferences in gambling. *American Journal of Psychology*, 1953, 66, 349-364.
- Eiferman, R. Response patterns and strategies in the dynamics of concept attainment behavior. *British Journal of Psychology*, 1965, 56, 217-222.
- Gentile, J., & Schippers, L. Differential pay-off-loss in probability learning and decision making. *Psychonomic Science*, 1966, 4(2), 65-66.
- Karlins, M., & Lee, R. Creativity and information search in a problem solving context. *Psychonomic science*, 1967, 8(4), 165-166.
- Wikoff, R. Personality correlates of psychometric response styles. *Dissertation Abstracts*, 1966, 27(1-B), 298.

REDUCTION OF AUTOMATED READABILITY INDEX

CALCULATION TIME

Joseph D. Young

Air Training Command

An automated readability index apparatus generated data for determining the reading difficulty levels of training materials. An evaluation of the apparatus revealed a loss of time in hand calculation of the index from machine-generated data. An experimental hand calculator proved efficient in significantly reducing computational time ($p < .05$)

The standard method of determining reading difficulty levels for training literature is the fog count. This method requires an individual to make a determination of the value of a given word received. To remove the judgmental aspect and automate the determination of readability levels, Smith & Senter (1967) developed a counting device connected to an electric typewriter. This device is called the automated readability index apparatus. They also devised a formula (Automated readability index) for converting the number of strokes, words and sentences recorded on the machine into a number representing the difficulty level. Kincaid et al., (1967) validated this concept in a later study.

Scharf (1969) demonstrated that by using the apparatus, a typist could determine the reading level of a given sample of material as competently as an editor using the fog count method. However, a great deal of the typist's time was expended in calculating the readability index from machine-recorded data, and calculation errors were common. Therefore, a hand calculator was developed to reduce calculation time and computational errors.

Method

Subjects

Eight female subjects were used. Each held clerical positions within the technical school, and selection was based on the fact that use of the calculator could be incorporated into their present jobs.

Equipment

The calculator (see Figure 1) consisted of two logarithmic scales placed at right angles to each other (a and b). Both scales had a

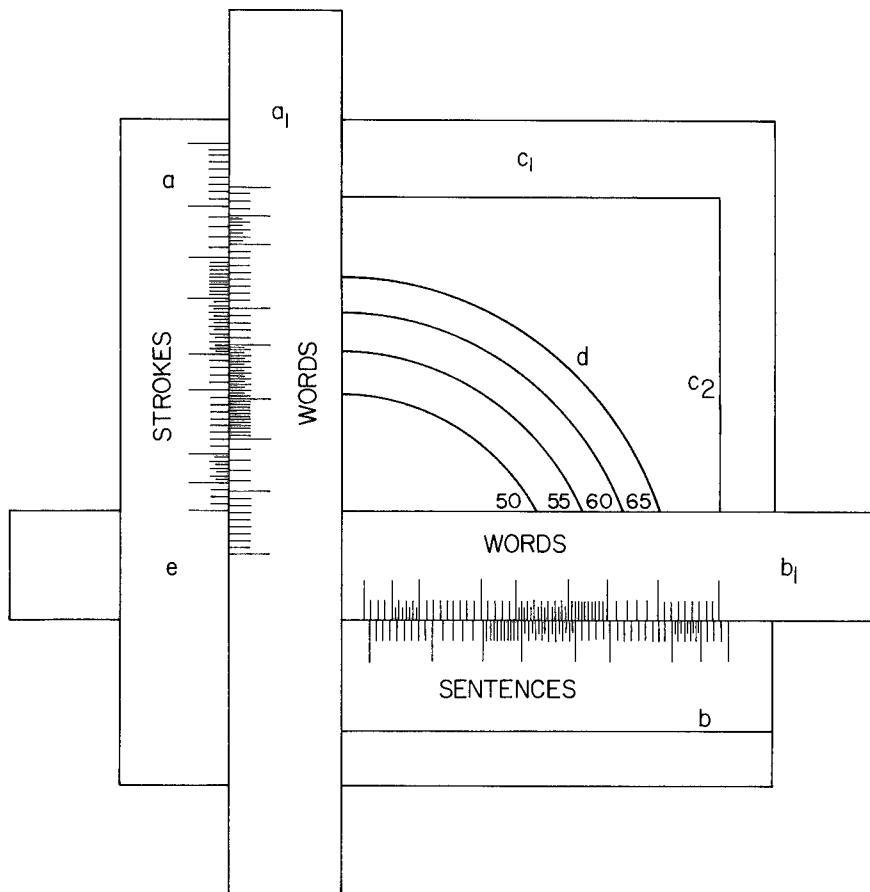


Figure 1
ARI HAND CALCULATOR

movable section (a_1 and b_1). In essence, these scales were identical to an ordinary slide rule in appearance and function. The only departure involved the addition of two course lines, one attached at each end of the a_1 and b_1 scale (c_1 and c_2 , other two not shown). As the a_1 and b_1 portions of the scales are moved back and forth, the course lines move over the logarithmic curves (d) numbered 50, 55, 60, and 65.

The Automated Readability Index consists of the following equation: $ARI = 9 (\text{strokes/words}) + (\text{words/sentences})$. The calculator solves the equation in the following manner: [1] 9 strokes/words is accomplished by the movement of scale a_1 relative to scale a , while multiplication by 9 is achieved through permanent positioning of the lower end of scale a relative to the lower end of scale b (area e); [2] (words/sentences) is calculated by the movement of scale b_1 relative to scale b ; and [3] summation is accomplished through the relative positions of course lines c_1 and c_2 .

In normal operation, a person matches the number of words on scale a_1 with the number of strokes on scale a , matches sentences on

scale b_1 with words on scale b , then notes the position on the graph (d) where course lines c_1 and c_2 cross. By noting the position of the crossed course lines, the individual can read the ARI directly from the logarithmic curves on the graph.

Only four number lines were used on the logarithmic graph. Since an ARI of 65 exceeds the desired reading level of eighth grade for technical training literature, any position above the 65 number line indicates the need for a revision of that material. Likewise, an indicated ARI of 50 or lower represents a reading level of fourth grade or less.

Procedure

Each subject was presented with a data sheet consisting of 20 data sets. Each data set had all information required to calculate an ARI equation.

Every subject was required to solve all 20 ARI equations, half by hand, while solving the other 10 problems with the ARI calculator. Four subjects used the calculator for the first 10 problems, while the other four subjects solved the first 10 sets by hand. The counterbalancing was used to diminish practice effects. Start and finish time was recorded for each subject's calculation by hand and by machine. In addition, an error allowance of ± 1 was used in checking the answers of the machine and hand calculated sections of each subject's data sheet.

Analysis consisted of comparing calculation time between the two methods, as well as total errors per subject between hand and machine calculation.

Results

TABLE 1

Mean Calculation Time in Minutes

Hand	Machine
18.88	9.50

Using a t test for correlated means, the reduction in calculating time proved to be highly significant ($t = 15.0$, $df = 7$, $p < .0005$).

TABLE 2

Mean Errors per Method

Hand	Machine
5.25	3.00

Likewise, reduction in calculation errors was significant ($t = 2.2$, $df = 7$, $p < .05$).

Discussion

Use of the ARI calculator appears to satisfy the need for reduced calculation time and increased accuracy. By eliminating the need for mental calculation, both speed and accuracy are improved. Moreover, both of these results appear to be direct benefits of employing the ARI calculator.

Although time savings were highly significant, error reduction was only significant at the 5 percent level. A combination of two factors contributed to this situation; namely, interpretation and parallax effect.

The subject was required to make a judgment if the course lines did not converge on one of the four number lines. By adding more logarithmic curves, the need for operator judgment would be diminished with the result of greater accuracy.

Parallax results from course lines on the calculator being two inches apart, one under the other. Consequently, a person can record the wrong ARI number if he does not look directly down on the crossed lines. Here again, by improved engineering, this can be eliminated. With the incorporation of these improvements, mean error could drop to 1.0 or less.

Conclusion

The demonstrated results of this calculator, with the performance of the ARI apparatus, promise to add a new dimension to the area of technical literature review. The ARI system (calculator and apparatus) makes it possible to arrive at a reading difficulty level while the rough draft is being typed. The difficulty index is less variable in interpretation than a fog count, more accurate, and done in a fraction

of the time. Consequently, all of this can be accomplished by a GS-3 typist, rather than the GS-12 editor usually needed for estimating reading difficulty levels.

References

- Guide for Air Force Writing*, AFP 10-1. U. S. Government Printing Office, Washington, D. C. 1969.
- Kincaid, J. P., Yasutake, J. Y., & Geiselhart, R. *Use of the automated readability index to improve comprehensibility of technical orders*. Systems Engineering Group, Technical Report 67-47, 1967.
- Scharf, G. P. *A feasibility test of the automated readability index apparatus*. Air Training Command, Project Report 69-22, 1969.
- Smith, E. A., & Senter, R. J. *Automated readability index*. Aerospace Medical Research Laboratory, Technical Report 66-220, 1966.

UTILIZATION, SUCCESS AND BENEFITS OF THE USAFI GED PROGRAM

William E. Beusse

Air Force Human Resources Laboratory

This study reports on factors associated with participation and success in the USAFI High School General Educational Development program for a group of first-term personnel who recently completed their active duty tours. It was determined that significant differences exist between participants and non-participants, achievers and non-achievers along such background characteristics as age at entry, race, educational level, aptitude level, source of accession, military service and military occupation.

The Department of Defense offers servicemen a wide range of educational benefits during their service careers. Existing programs include high school, college and graduate school attendance, correspondence courses, self-study courses, and high school and one year of college equivalency examinations. These programs are designed to provide opportunity for military personnel to acquire knowledge and skill to assist them in personal as well as occupational growth. It is felt that they produce a more productive serviceman both in his military job and in his civilian life when he leaves the military service.

One of the largest of the Department of Defense educational programs in the High School General Educational Development Tests (GED) which are administered by the United States Armed Forces Institute (USAFI) is Madison, Wisconsin. These tests measure the extent to which an individual has acquired the equivalent of a general high school education. The original forms of these tests were developed during World War II and were based on the philosophy that it should be possible to measure the knowledge an adult acquires informally and compare this informal educational level with that of the educational level of an individual who had obtained his education through formal schooling.

The core GED curriculum offers nearly 100 typically high school subjects including English, mathematics, social studies, sciences, and business subjects. However, the courses are not a prerequisite for taking the GED tests.

The data for the study of utilization and success rates were extracted from the USAFI Student Master File and the DOD Post-Service Information File. The population consisted of 231,973 one-term

personnel (in all branches of the service) who entered the service as non-high school graduates, completed their active duty tour and separated from the Armed Forces during the period July 1968 through December 1969.

Utilization of the USAFI GED Program

The participation rate (the percentage of non-high school graduates who attempt to gain high school equivalency through the GED program) is taken as the measure of utilization. Of the 231,973 one-term servicemen, 59.4% (137,792) participated in the GED program.

Participation rates were found to vary by branch of service, race, source of entry into service, duty in Vietnam, marital status, and military occupation (using one-digit DOD occupation code).

The participation rates for various population breakouts are shown in Table 1. The results of this analysis indicate that:

The more technically oriented services, the Air Force and Navy, have the highest participation rates.

Enlistees are far more likely to participate than inductees.

Negroes are less likely to participate than Caucasians.

Men who served in Vietnam are less likely to participate than those who did not.

Married men are slightly more likely to participate than single men.

Men in the higher skilled military occupations (i.e., electronics, technical, etc.) are more likely to participate than men in the lower skilled military occupations (i.e., infantry, service and supply).

In addition, it was found that there is a direct relationship between aptitude, as measured by the Armed Forces Qualification Test (AFQT), and participation (See Table 2). The higher a serviceman's aptitude, the more likely he is to participate. This strong relationship between AFQT and participation seems to explain the differences between Negro and Caucasian participation rates. When AFQT is held constant (within decile ranges) the differences disappear. In fact, at least in the low AFQT ranges (scores 10-49) Negroes have higher participation rates.

There is also a direct relationship between educational level and participation. The higher a serviceman's level of civilian schooling, the more likely he is to take the GED tests.

TABLE 1

Participation and Achievement Rates

Population Breakout	Participation Rates (%)	Achievement Rates (%)
Branch of Service		
Air Force	87.8	71.3
Navy	74.0	58.2
Marine Corps	58.7	58.7
Army	54.4	65.2
Race		
Caucasian	60.4	64.9
Negro	57.7	48.9
Source of Entry		
Enlisted	69.3	65.2
Inducted	44.7	58.6
Duty in Vietnam		
Yes	54.1	63.2
No	65.2	63.1
Marital Status		
Single	58.7	62.5
Married	60.8	64.5
Military Occupation		
Electronic Equipment Repair	71.7	73.7
Other Technical Specs.	69.1	72.3
Medical & Dental Specs.	68.9	78.3
Admin Specs. & Clerks	68.4	73.0
Communication & Intelligence	64.5	70.1
Elec/Mech Equip. Repair	63.5	64.3
Craftsmen	60.7	62.5
Infantry, Gen. Crews, Seamen	53.8	56.2
Service & Supply	51.6	58.2

TABLE 2

Mean Characteristics of Participants and Non-participants

Variable	Partici- pants	Non- Partici- pants	Point Biserial r	Significance Level
Age at Entry	18.79	19.71	.23	.01
Highest Year of Education Completed	9.86	9.64	.10	.01
AFQT	45.76	32.48	.30	.01
Pay Grade at Separation	4.04	3.89	.08	.05

2

TABLE 3

Mean Characteristics of Achievers and Non-achievers

Variable	Achievers	Non- Achievers	Biserial r	Significance Level
Age at Entry	18.70	18.96	.09	.05
Highest Year of Education Completed	10.00	9.68	.18	.01
AFQT	50.82	36.98	.39	.01
Pay Grade at Separation	4.10	3.93	.11	.01

Pay grade is also positively related to participation. Servicemen who left service at the higher pay grades were more likely to have participated in the GED program.

There is an inverse relationship between age at entry and participation. The younger a serviceman is when he enters the military, the more likely he is to participate.

Success in the USAFI GED Program

The achievement rates for various population breakouts appear in Table 1. The results of this analysis indicate that:

The Air Force and Army have the highest achievement rates.

Enlistees are more likely to pass than inductees.

Negroes have a lower achievement rate than Caucasians.

Service in Vietnam does not have any effect on achievement rates of GED program participants.

Married men are slightly more likely to pass the GED tests than single men.

Men in the higher skilled military occupations (i.e., electronics, medical and dental) are more likely to pass than men in the lower skilled military occupations (i.e., infantry, service and supply).

There is a direct relationship between AFQT and achievement. The higher a serviceman's aptitude the more likely he is to pass the GED tests. It is this relationship between AFQT and achievement that explains the differences between Negro and Caucasian achievement rates. When AFQT is held constant (within decile ranges) the differences disappear.

There is also a positive relationship between educational level and achievement. The higher a serviceman's educational level, the greater the likelihood that he will pass.

Pay grade is also positively related to achievement. Servicemen who left the service at the higher pay grades were more likely to successfully complete the GED program.

Age at entry is negatively related to achievement. The younger the serviceman, the more likely he is to pass the GED examinations.

Benefits of the GED Program

The data for this part of the study were obtained by means of a mail survey. Questionnaires were sent to a stratified random sample of 4000 consisting of 1000 from each of the following four groups: Those who passed the GED tests at the level required by their state; those who passed at the DOD recommended level; those who participated and failed; and those who did not participate.

Preliminary analyses of the initial returns indicate that certain benefits accrue to men who successfully complete the GED program.

71.4% of the respondents state that having the USAFI certificate has helped them in civilian life.

Of those men who tried to enter educational or training programs, 74.4% stated that their USAFI certificates were accepted as evidence of high school completion.

It was also found that the better a man does on the GED tests, the more likely he is to continue his education and utilize the educational benefits of the GI Bill.

In conclusion, this research indicated that tangible benefits accrue to those men who successfully complete the GED high school equivalency program. The task remains, however, to find ways to bring the program to those groups (low AFQT, low educational level, etc.) who are least likely to participate and pass the GED examinations.

Drugs and Rehabilitation

CHAIRMAN: Captain Donnell L. Washington

THE VIETNAM HEROIN EPIDEMIC:

A DESCRIPTIVE PROFILE OF THE ARMY RETURNEE

Larry H. Ingraham

Walter Reed Army Institute of Research

Army enlisted men who had been identified as heroin users upon leaving Vietnam were interviewed to assess their perceived need and desire for further rehabilitative treatment. Demographic details, drug use histories, evaluations of the detection and detoxification programs, and estimates of future drug use were obtained. Group interviews were used to explore the social conditions integral to heroin use in Vietnam. Results of the interviews are summarized, and suggestions for needed future research are presented.

Over the past 18 months observers from the Defense Department, the Congress, and the media have agreed that heroin usage by American troops in Vietnam reached epidemic proportions which posed a threat to both the defense capability and to the national health. The DOD counter-offensive on drug abuse launched in mid-1971 required in part: (a) mandatory urine tests for opiate derivatives for all military personnel leaving Vietnam, (b) mandatory detoxification for those found to be opiate-positive, and (c) mandatory participation in drug rehabilitation programs at either active duty military or VA installations.

The objectives of the present research were three: (a) to provide a profile of the demography, social history, and drug use patterns of the returning veterans, (b) to provide a preliminary assessment of the impact of mandatory detection and detoxification on the expressed need and desire for future rehabilitative opportunities, and (c) to provide strategies and hypotheses to guide our future research efforts.

Method

In pursuit of these objectives, 78 opiate-positive returnees were interviewed at four different Army posts in CONUS during the final three months of 1971. Army enlisted psychology/social work technicians conducted individual structured interviews, and a social psychologist, an Army Captain, conducted group interviews that were essentially unstructured.

Results

Four major conclusions are warranted by the data. The first major conclusion is that, as a group, the opiate-positive returnees are not markedly different from opiate-negative returnees in the enlisted ranks. Both the mean age in our sample (21 years) and the ethnic proportions (67% white, 24% black) closely approximated the population values. Disproportionate representation of urban backgrounds, broken homes, and low educational levels was *not* evident in the sample, and nearly three-fourths of the respondents reported no civilian police record.

Similarly, the military records of the respondents also appear to be typical of the general enlisted population. Three-fourths of the respondents were volunteers, and over 90% evidenced normal progression in rank. The majority (56%) reported no record of disciplinary action related to drug use.

The second major conclusion is that the opiate-positive Vietnam returnees, as a group, had considerable experience with illegal drugs prior to entering military service. Excluding under-age alcohol usage, 54% reported experience with at least one illegal drug regularly on at least a weekly basis. However, 81% of the respondents had not tried heroin before they went to Vietnam, and 28% had tried no illegal drugs whatsoever.

The third major conclusion is that the respondents emphatically deny both the need and the desire for further rehabilitation opportunities. This denial does not appear to be related to the treatment they received in either the detection or detoxification program in Vietnam. The delay in returning to friends and families, the lack of authoritative information concerning their future treatment, and their being incarcerated like prisoners or mental patients angered and frustrated the respondents. However, no one objected, in principle, to having to submit a urine sample or having to undergo detoxification before leaving Vietnam. For our respondents, detoxification was considered the treatment. Since they had been already detoxified at the time of our interview, our questions of their need for further treatment or rehabilitation were regarded as irrelevant and utter nonsense.

This leads to the fourth major conclusion and to the thesis of this presentation. In order to understand the Vietnam heroin epidemic, it is necessary to carefully consider the structure and dynamics of the social system in which heroin is used in Vietnam. From our interviews it is very apparent that drug usage is not primarily an end in itself for the vast majority of the users. We are not so much dealing with individual heroin users but with a well-structured, albeit informal, primary social group phenomenon. Drug

use provides a means for securing distinctive kinds of social experience and to insure these experiences, a primary social system (actually a sub-system within the larger social context) has evolved whose members call themselves "heads."

Common preferences for music, art, dress style, and hair length are markers of the "head" social system. Within the sub-system, a distinctive language pattern permits members to communicate with each other and to distinguish other members from outsiders. A set of behavioral rules regulates the system by defining how the "heads" are expected to treat each other and how they are to behave toward non-members. Finally, the social system of "heads" is marked by an ideology that sharply distinguishes it from other primary social groups whose members are not "heads."

The sub-system is designed to provide a variety of social experience with minimal investment in the acquaintance process or in interpersonal commitment. When an individual becomes a "head," he moves freely throughout the living quarters of many men and has access to a variety of entertainments available on radios, tape decks and televisions. He enjoys conversation and companionship that he describes as honest, open, and deeply meaningful. The conjunction of drug effects with novel auditory and visual sensations provided by hard rock music, black lights, and psychedelic art provide a common experiential referent for conversation. Relationships are described as quiet, calm, without acrimony or braggadocio, and without fear of criticism or censure. The "head" also shares food, money, possessions, and drugs with other members of the system, and he describes the acts of sharing as very gratifying. These social experiences are available with minimal acquaintance and interpersonal commitment. Further, the "head" can readily establish his membership credentials in the "head" sub-system at almost any military unit at any geographic location in Vietnam.

Although not often expressed directly by our respondents, another benefit of becoming a "head" is protection from theft and assault. Strong normative expectations were expressed that "heads" do not steal from each other. Not only is there a system rule that prohibits theft, but also the relatively free access to others' possessions increases surveillance opportunities which would decrease the likelihood of successful theft by an outsider.

The "head" protects himself from assault in two ways. First, his group activities are conducted either in private quarters or out of the mainstream of traffic in the unit. Relative isolation coupled with the definite rule that prohibits violence among "heads," decreases the probability of encountering bellicose individuals. Secondly, in the event of assault, the cohesive nature of the "head" system assures support in fending off assailants or in providing revenge.

The system benefits of social interaction and protection are available at minimum cost. To become a "head," an individual need only fail to condemn drug usage by others and actively participate in the social activities of the "head" system. It must be emphasized however, that drug use, *per se*, is neither a sufficient nor a necessary condition for inclusion in the social system described.

Membership is determined by willingness to endorse the values and rules of the primary group and not by the amount or kind of drugs used. Even for members who use drugs, little significance is attached to whether a man uses marijuana, barbiturates, amphetamines, or heroin. In short, the heroin user enjoys no special status within the "head" system.

Within the social system as it functions in Vietnam, heroin users do not consider themselves "addicts" or "street junkies" in the same connotative sense as the terms are used in the civilian sector. Hence, the use of heroin is an adjunctive activity rather than a role with self-attributed status. By and large, they retain an abhorrence of the dope-fiend addict portrayed in the civilian stereotype. They grant that the stereotype junkie with a long record of crime, drug use, and dereliction needs treatment and rehabilitation, and to the extent their members exhibit these characteristics, they exclude them from the social system. However, most of the returnees had not behaved in ways consistent with their stereotype of "street junkies," hence they saw no reason for treatment beyond detoxification.

Discussion

In closing, I want to speculate briefly on three policy factors that may contribute to the emergence and maintenance of the "head" social system.

The first factor is the 12-month rotation policy which has made it difficult for the individual soldier to identify with his unit as a primary social group. In our interviews, the collective "we" always referred to "heads" and not to the local platoon or company. The continuous rotation of new men and leaders in the units has vitiated the first principle of military leadership, "Know Your Men;" hence the unit has been reduced to a block on the organization charts that has administrative but not psychological significance for the individual.

The second factor is the style of warfare necessitated by the conditions in Vietnam. The respondents repeatedly spoke of the frustration of not knowing who the enemy was and of having to undertake missions that lacked meaningful objectives. In Vietnam during the retrograde action, there have been few bunkers to charge, bridges to capture, or critical terrain to hold in the legendary John Wayne

fashion. These factors coupled with mounting anti-war sentiments in both the civilian and military sectors seem to have left the individual soldier with a profound sense of futility concerning his contribution to the military effort.

The third factor is the high degree of affluence within the American Army. The "head" sub-system has already been described as a stable group that cuts across units and geography. Within the system, psychological identification is possible, and material possessions provide the means for achieving social status. During the draw-down of the war, my respondents reported working regular eight-hour days both in the rear and at the fire bases. In addition to considerable leisure time, they also enjoyed the affluence of stereos, tape decks, black lights, poster art, and televisions while in the war zone. Within the "head" system, those with the best or newest equipment are accorded the highest status and thus provide the focus for social interaction within the system. An individual who must sell his possessions to support his heroin habit is devalued and extruded from the social system.

Stripped of the possibility of psychological identification with the military unit and denied the opportunity of securing social worth by personal risk-taking for socially approved objectives, it is small wonder that a social system emerged that provided both stability and individual self-enhancement. In addition, the social system provides a means of dealing with tension and depression through drug use, as well as an ideology that is diametrically opposed to the perceived doctrine of military violence.

In conclusion, the most significant aspect of the Vietnam heroin epidemic is the existence of a primary social group which encourages and maintains drug use. The existence of the sub-system in Vietnam is unquestionable, but the scope of such a sub-system within the military and the organizational factors that contribute to its emergence and maintenance are at present only speculative. The greatest potential for both significant social research and rational intervention lies in our understanding the structure and dynamics of the "head" system as it may exist in minor transformations throughout the military and in identifying the organizational factors that account for its emergence and continued functioning.

Footnote

¹I wish to express my most sincere appreciation to David H. Marlowe, Ph.D., Walter Reed Army Institute of Research, for his kind encouragement and skillful criticism during the preparation of this manuscript. Thanks are also due to Christine Yowell, Robert Matthews, and Mark Gutwein who assisted in collecting and analyzing the data.

COMPARISON OF PERSONAL CHARACTERISTICS OF
IDENTIFIED DRUG USERS WITH NON-USERS

Bart M. Vitola and Cecil J. Mullins

Air Force Human Resources Laboratory

Comparisons were made between heroin users, control groups, and the total USAF Vietnam population on test performance, age, education, race, and Air Force Specialty variables. When compared to their control groups, heroin users are less well educated, show less ability, cluster in the 19 and below through 21-year old age group, and appear to be found in significant disproportion in particular Air Force Specialties. These findings partially characterize the potential heroin user and imply the need for more in-depth recruiting interviews and greater intensification of existing drug education programs for particular age groups and Air Force Specialties.

For over a decade, stringently-worded drug laws have been passed by United States legislators in the hope of bringing about a reduction in the incidence of narcotic addiction; narcotics being defined as opiate derivatives or their synthetic equivalents (opium, morphine, heroin, meperidine, methadone, and codeine). According to Louria (1968), an attempt to legislate away the narcotics problem will be about as successful as was the attempt to legislatively control the alcohol problem of the 1920s. In his book entitled, *The Drug Scene*, Louria states that even increased awareness of the severity of the heroin problem and the strongly-worded laws passed in the 1950s have failed to bring any reduction in the incidence of narcotic addiction in the United States over the last fifteen years.

Evidence gathered from the last ten years of research has provided a profile of the average heroin addict, Chein (1964), Einstein (1966), Kron (1965), Laskowitz (1961), and O'Donnel (1967). Generally speaking, the addict is depicted as being rather young, not too well educated, offering minimal skills to the labor market, and about 68% of the time, coming from a repressed minority. Noteworthy is the fact that few, if any, studies characterizing the heroin addict include a control group of the same age, exposed to the same environment, having the same educational background or approximately the same average skills. The goal of this study is to characterize the heroin user as to age, aptitude and educational level, and race by comparing groups of heroin users and their control groups.

Method

Data were collected from the Uniform Airman Record (UAR) and the Air Force Drug Research Data Base file for the time period July through October 17, 1971, on three groups of enlisted men in Vietnam. Sample 1 was comprised of 296 heroin-users, identified through urinalysis and the Limited Privilege Communications Program. Sample 2 contained 888 control subjects. Samples 1 and 2 were matched on age, duty location, and date of return. The third group was comprised of 31,815 men who were the total enlisted Air Force population in Vietnam as of June 30, 1971. It is possible than unidentified users were in the control sample and the total Vietnam group. Therefore, differences between user and comparison groups would tend to be conservative.

Distributions were obtained comparing Samples 1 and 2 on age, racial subgroup, educational background and Airman Qualifying Examination and Armed Forces Qualification Test performance. In addition, Air Force Specialty (AFS) information was obtained indicating the proportion of the total Vietnam population (N = 31,815) that each AFS represented and the incidence of drug use within the AFS.

Results and Discussion

In analyzing the data, the first fact which became evident was the abuse rate. Of the 31,815 enlisted men in Vietnam as of June 30, 1971, Air Force identified 296 as heroin users. These data indicate a less than one percent abuse rate for the time period of July through October 17, 1971. Table 1 presents distributions which show percentages of identified drug users by racial subgroup.

Considering the Negro-non-Negro representation in Vietnam (14 percent versus 86 percent), the data suggest that the amount of heroin abuse among Negroes in Vietnam is disproportionately high. However, it should be noted that the incidence of heroin abuse among Negroes as a racial subgroup in the civilian community is high. For example, in New York State, which accounts for one-half of all known heroin addicts in the United States, 50.4 percent are Negro, 13.6 percent are Puerto Rican, and 5.4 percent are Mexican (Louria, 1968). Based on these data, it appears that heroin addiction is, quantitatively speaking, a disease of repressed minorities.

Heroin Abuse by Educational Level

To indicate heroin abuse by education level for users and their control groups, distributions for years of formal education completed by racial subgroup were obtained and are presented in Tables 2 and 3. When levels of education of heroin users were compared with those of their control groups and with those of the total Vietnam population,

TABLE 1

Distribution of Heroin Users in Vietnam
by Racial Subgroup

Racial Subgroup	Number and Percentage for Racial Subgroup			
	User		Total Vietnam Population	
	N	%	N	%
Negro	157	53	4,425	14
Non-Negro	139	47	27,390	86
Total	296	100	31,815	100

TABLE 2

Distribution of Heroin Users and Control Groups in Vietnam
for Various Levels of Education by Racial Subgroup

Years Schooling Completed	Number and Percentage for Educational Level											
	Negro (N=277)				Non-Negro (N=907)				Both Groups Combined (N=1,184)			
	User		Control Group		User		Control Group		User		Control Group	
	N	%	N	%	N	%	N	%	N	%	N	%
16 or more	0	.00	2	1.67	1	.72	13	1.69	1	.34	15	1.69
13-15	3	1.91	3	2.50	6	4.32	37	4.82	9	3.04	40	4.50
12	124	78.98	109	90.83	111	79.86	672	87.50	235	79.39	781	87.95
11 or less	30	19.11	6	5.00	21	15.10	46	5.99	51	17.23	52	5.86
Total	157	100.00	120	100.00	139	100.00	768	100.00	296	100.00	888	100.00

TABLE 3

Distribution of Total Vietnam Population
for Various Education Levels by Racial Subgroup

Years Schooling Completed	Number and Percentage for Educational Level					
	Negro		Non-Negro		Total Group	
	N	%	N	%	N	%
16 or more	35	.79	741	2.71	776	2.44
13-15	244	5.51	2,122	7.75	2,366	7.44
12	3,703	83.68	22,058	80.53	25,761	80.98
11 or less	443	10.02	2,469	9.01	2,912	9.14
Total	4,425	100.00	27,390	100.00	31,815	100.00

two facts obtained significance. Regardless of race, heroin users as a group are less educated than their comparison groups and there is, across racial groups a disproportionate number of users who have not completed high school. The last fact may become of primary significance when determinations are being made concerning the minimum level of education required of potential enlistees for an all-volunteer force.

Heroin Abuse by Age Group

A research of the literature concerning the age range in which heroin abuse is most prevalent leads one, after much reading, to the conclusion that the average age of users will vary with geographic location and the availability of the drug. Tables 4 and 5 present distributions by age and racial subgroup for heroin users in Vietnam.

TABLE 4

Distribution for Heroin Users in Vietnam by Age and Racial Subgroup

Age Group	Number and Percentage for Age Group					
	Negro		Non-Negro		Both Groups	
	N	%	N	%	N	%
19 and below	10	6.37	19	13.67	29	9.80
20 years	41	26.11	38	27.34	79	26.68
21 years	43	27.39	42	30.22	85	28.72
22 years	35	22.29	21	15.11	56	18.92
23 years	19	12.11	11	7.90	30	10.14
24 and over	9	5.73	8	5.76	17	5.74
Total	157	100.00	139	100.00	296	100.00

TABLE 5

Distribution for Total Vietnam Population by Age and Racial Subgroup

Age Group	Number and Percentage for Age Group					
	Negro		Non-Negro		Both Groups	
	N	%	N	%	N	%
19 and below	218	4.92	1,102	4.02	1,320	4.15
20 years	506	11.44	2,829	10.33	3,335	10.48
21 years	666	15.06	4,411	16.11	5,077	15.96
22 years	652	14.74	3,844	14.03	4,496	14.13
23 years	329	7.43	2,161	7.89	2,490	7.83
24 and over	2,054	46.41	13,043	47.62	15,097	47.45
Total	4,425	100.00	27,390	100.00	31,815	100.00

It appears that with the exception of the 22-year old Negro, the same patterns of heroin use by age group are found within racial subgroups. When the total heroin-user sample is compared by age levels to the age levels of the total Vietnam population, it becomes apparent that the ages 19 and below through 21 are those of greatest vulnerability. This group represents 31 percent of the total Vietnam population and 65 percent of the heroin-user population. The 22 and 23-year old age groups represent 22 percent of the Vietnam population and 29 percent of the heroin-user population. The 24-year olds and above represent 47 percent of the Vietnam force and 6 percent of the heroin-user population. These data indicate an inverse relationship between age and extent of participation in heroin use.

Test Performance Related to Drug Use

Distributions of AFQT scores and AQE aptitude indexes for user and control groups were obtained to determine the relationship between test performance and drug use. The results are shown in Tables 6 and 7. In summary, the data indicate that, across racial subgroups, there are moderate differences between mean AFQT and AQE scores on users and their control groups; users consistently score lower.

When AFQT and AQE scores of the total user group are compared to those of the total Vietnam population, the mean differences become more dramatic. The differences in AFQT scores is 16 centile points. AQE mean differences ranged from six to nine centile points.

It appears, in addition to previous characterization, the heroin user may be described as having less ability to learn than his peers.

TABLE 6

Aptitude Index and AFQT Performance for Heroin Users and Control Groups in Vietnam by Racial Subgroup

Test Measure	Negro				Non-Negro				Both Groups Combined			
	User (N=157)		Control (N=120)		User (N=139)		Control Group (N=768)		User (N=296)		Control Group (N=888)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
AFQT	30.55	15.75	34.32	18.78	53.94	23.99	58.95	23.04	41.50	23.19	55.68	24.02
Mechanical AI	45.71	17.10	48.54	19.70	57.66	19.54	59.79	19.82	51.34	19.24	57.74	20.50
Administrative AI	46.43	19.53	52.52	18.93	57.84	19.43	61.60	19.40	51.82	20.30	60.38	19.58
General AI	49.77	16.90	51.90	17.23	60.90	16.94	63.46	17.35	55.03	17.82	61.91	17.78
Electronics AI	43.81	17.85	46.40	19.33	57.95	18.36	62.57	19.11	50.47	19.42	60.40	19.92

TABLE 7

Aptitude Index and AFQT Performance for Total Vietnam Population
by Racial Subgroup

Test Measure	Negro (N=4,425)		Non-Negro (N=27,390)		Total Population (N=31,815)	
	Mean	SD	Mean	SD	Mean	SD
AFQT	34.80	18.43	60.75	23.10	57.10	24.24
Mechanical AI	42.57	19.68	61.10	20.92	58.52	21.72
Administrative AI	48.77	19.91	61.79	21.15	59.98	21.46
General AI	49.82	17.63	63.32	19.47	61.23	19.92
Electronics AI	45.05	19.05	61.89	21.20	59.38	21.80

2

Drug Abuse by Air Force Specialty

To identify those Air Force Specialties in which a significant amount of heroin use was taking place, distributions of the total Vietnam sample and racial subgroup were obtained. Of the 47 AFSs in Vietnam as of June 30, 1971, 19 showed no incidence of drug abuse, represented 6 percent of the total population, and in most instances required an entrance AQE minimum of 60 or above. Fourteen AFSs showed little or moderate abuse and represented 48 percent of the total Vietnam population. Fourteen AFSs showed a disproportionate amount of drug use, represented 46 percent of the total Vietnam population, and in most instances required an entrance AQE minimum below 60.

It is to these last 14 AFSs that the data of Table 8 address themselves. Although all of the 14 AFSs represent a potential drug abuse problem, Guilford's test to determine the significance between heroin and control group proportions of the total Vietnam population was applied to each of the 14 specialties (Guilford, 1965). The results showed that in eight specialties there was a significantly high incidence of drug use. These eight AFSs represented 36 percent of the total Vietnam population and 59 percent of the total drug-user population. Of special note is the fact that entrance into all of these AFSs could be obtained with a minimum AQE aptitude index of 40.

Conclusions

In conclusion, the following statements and suggestions may be made concerning the characteristics of heroin users:

1. The ages 19 and below through 21 appear to be the ages of greatest heroin-user vulnerability.

TABLE 8

Significance of Differences Between Proportions of Heroin Users
and Comparison Groups for Specific
Air Force Specialties in Vietnam

Specialty Code	Entry AI	Air Force Specialty	Total Population in Vietnam			P
			Users (N=296)	Comparison Group (N=31,519)	Both Groups Combined (N=31,815)	
40	ME-40 E-60	Intricate Equipment	.68	.24	.25	ns
42	ME-40 ME-50	Aircraft Accessories	7.09	4.59	4.62	.05
47	M-40	Vehicle Maintenance	2.36	1.82	1.83	ns
53	M-40 MG-50	Metalworking	2.03	1.62	1.63	ns
55	MAG-E-40 M-50 A-60 G-65	Civil Engineering Structural Pavements	4.73	2.75	2.77	.05
57	G-40	Fire Protection	4.39	1.12	1.15	.01
60	MAG-40 MA-50	Transportation	11.50	6.47	6.52	.01
62	G-40	Food Services	3.04	1.59	1.62	.05
63	MGE-40	Fuel Services	2.36	1.25	1.26	.05
64	G-40 AG-60 A-70	Supply	7.09	6.33	6.33	ns
70	A-40 AG-60	Administration	11.82	7.74	7.77	.01
71	G-40	Printing	.34	.07	.08	ns
81	G-40	Security Police	14.29	10.48	10.51	.05
98	G-60	Dental	.34	.25	.25	ns

2. When compared to non-Negroes, the rate of heroin use among Negroes is extremely high (53% drug versus 14% Vietnam).

3. Regardless of race, heroin users as a group are less well educated and display lower aptitude levels than their comparison groups.

4. There should be an intensification of programs dealing with the consequences of drug use in particular AFSs, especially among first term airmen.

5. Prior to enlisting a high school non-graduate who demonstrates a low aptitude potential, recruiters should make every effort to determine the extent of the applicant's involvement in drug usage.

References

- Chein, I., Gerard, D., Lee, R., & Rosenfeld, E. *The road to H: Narcotics, delinquency and social policy*. New York: McGraw-Hill, 1965.
- Einstein, S. The narcotics dilemma: Who is listening to what? *International Journal of Addictions*, 1966, 1(2), 2-6.
- Guilford, J. P. *Fundamental statistics in psychology and education*. New York: McGraw-Hill, 1965.
- Kron, Y. J., & Brown, E. M. *Mainline to nowhere*. New York: Pantheon, 1965.
- Laskowitz, D. The adolescent drug addict: An Adlerian view. *Journal of Individual Psychology*, 1961, 17, 68-79.
- Louria, D. B. *The drug scene*. New York: McGraw-Hill, 1968.
- O'Donnel, J. A., Besteman, K. J., & Jones, J. P. Marital history of narcotics addicts. *International Journal of Addictions*, 1967, 2(1), 21-39.

ATTITUDE CHANGE AMONG SELECTED AIR FORCE PRISONERS

Victor N. Ogilvie

Air Training Command

The study tested the effectiveness of the 3320th Retraining Group in preparing selected Air Force prisoners for return to active duty by promoting positive attitudes. Sixty male, USAF prisoners, participating in a rehabilitation program, were administered a battery of tests which provided scores on attitudes about themselves, others, obeying the law, patriotism, and the Air Force. While the results yielded little evidence that attitudes became more positive, it was revealed that successful retrainees had higher initial mean scores on three scales than the failures, and maintained this superiority throughout the duration of the entire program.

Since its inception in 1951, the primary mission of the 3320th Retraining Group has remained relatively unchanged, and one of its main roles, according to Air Force Manual 125-2, has been "to prepare retrainees for return to active duty by encouraging in them a wholesome and favorable attitude toward their immediate environment, the Air Force, and society." During this time the Program Evaluation (research) Division of the Retraining Group has conducted numerous empirical and descriptive studies on the characteristics of the retrainee population, the nature of the rehabilitative process, and the assessment of program results. But if one thing has been made apparent, it is that this body of research is characterized by a relative dearth of information on retrainee attitudes and the extent to which they do change and in what direction.

Other research studies dealing with attitude change among delinquent adolescents and prison inmates (Deitz, 1970; Hamner, 1969, Brown, 1970; Kelly & Baer, 1969; Aitken, 1969; and Gattshall, 1969) are characterized by generally inconclusive findings.

In spite of this compilation of research data, virtually no information is available with which to gauge the growth of positive attitudes held by the retrainee about himself, other people, the Air Force, and society in general. To be sure, retrainee attitudes do change some during the period of confinement and rehabilitation, but do they necessarily become more positive? This is the question that has not yet been satisfactorily answered.

The following hypotheses were tested:

1. Change in expressed attitudes will be significantly greater among airmen in the retraining program than among airmen in a technical school course of instruction.
2. Attitude change will be significantly greater among those retrainees returned to duty than among those who fail to successfully complete the program and who receive a discharge under less-than-honorable conditions.

Method

Subjects

The subjects were 60 male Air Force prisoners participating in a rehabilitation program at the 3320th Retraining Group, Lowry Air Force Base, Colorado, during the early part of 1971. In order to insure heterogeneity of the sample, all retrainees who arrived at the Retraining Group during a 50-day period were selected for the study. An additional group of 65 male USAF technical school students served as a control group. These men neither participated in the formal retraining program, nor were they in confinement status. They represented four different technical specialties and were matched with the retrainees as closely as possible on the variables of age, intelligence, and length of military service.

Instruments

Four dependent variables were employed to measure the effects of the retraining program upon attitude change: (a) pre- and posttest scores on a scale measuring attitudes toward the Air Force (Remmers, 1960), (b) Pre- and posttest scores on a scale measuring attitudes toward obeying the law (Remmers, 1960), (c) pre- and posttest scores on a patriotism scale (Thurstone, 1932), and (d) pre- and posttest scores on a scale measuring acceptance of self and others (Berger, 1952).

Procedure

The scales were administered as part of a routine test battery to all Ss within two weeks of their arrival at the 3320th Retraining Group. During the orientation phase, each retrainee was randomly assigned to one of four treatment teams and remained there until his minimum release date (MRD) was reached and he was no longer in confinement. (Some men, however, did remain with the Group for a short period of time after their MRD in order to complete vocational training or to await final disposition of their cases.) Once the MRD was reached, retrainees were individually scheduled to take another form

of the same scales. Technical school students were tested during the first week of their respective classes and then again during the final week of each class. They were told that this was a research project which was concerned with their opinions on a number of subjects, and that their scores on these questionnaires would in no way influence their course grades. The retrainees were given similar instructions but were told that their scores would in no way influence their progress in the retraining program.

Results

In comparing scores of retrainees with those of USAF technical school students at the beginning and end of their respective programs, means and standard deviations were computed for each of the four treatment groups and are presented in Table 1. A two-way analysis of variance determined if any significant differences existed between the means of these groups. The analysis of variance data are presented in Table 2 and indicate several significant F values.

A Duncan's multiple-range test was performed to determine which specific groups actually differed significantly. While no significant main or interaction effects were revealed between the two groups at either phase of their training on the scales measuring attitudes toward others, numerous instances of significant differences between treatment means were found on the scales of patriotism, obeying the law, and the Air Force.

Table 3 shows means and standard deviations of scores of successful and unsuccessful retrainees at three different stages during their rehabilitation program. A two-way analysis of variance was used to determine if significant differences existed between the means of the groups. Table 4 is a summary of the five separate analyses of variance (one for each attitude scale).

A Duncan's multiple-range test was used to determine significant differences between the various means. Those retrainees who did not successfully complete their rehabilitation programs (i.e., failures) had a significantly higher mean score on attitudes toward self at the program's termination than did those men who successfully completed the program. On attitudes toward patriotism, obeying the law, and the Air Force a consistently higher mean score was found throughout all three phases of the program for the successful retrainees than for the failures.

TABLE 1

Mean Scores and Standard Deviations of Retrainees and Students on Attitude Scales

SCALES	BEGIN PROGRAM				END PROGRAM			
	Retrainees		Students		Retrainees		Students	
	M	SD	M	SD	M	SD	M	SD
SELF	128.15	26.70	137.60	15.48	136.80	25.84	136.54	16.30
OTHERS	102.43	11.36	102.97	10.39	101.57	11.66	100.00	10.76
PATRIOTISM	5.97	1.58	6.90	1.71	6.00	1.50	6.29	1.21
OBEYING LAW	7.55	1.16	7.65	.876	7.02	1.87	8.46	.768
AIR FORCE	5.70	2.29	6.34	1.82	5.40	2.11	6.56	1.83

TABLE 2

Two-way Analysis of Variance of Retrainee and Student Scores on Attitude Scales

SCALES	MS ERROR (df = 246)	MS TIME (A) (df = 1)	F	MS GROUP (B) (df = 1)	F	MS A X B (df = 1)	F
SELF	500.67	810	1.62	1317	2.63	1469	2.93
OTHER	76.35	448.89	5.879*	1.95	.025	189.72	2.484
PATRIOTISM	.662	5.70	8.61**	23.06	34.83	6.26	9.46**
OBEYING LAW	1.52	2.77	1.82	37.6	24.74	26.87	17.68***
AIR FORCE	4.05	.06	.01	47.5	11.73	7.31	1.80

*P<.025

**P<.005

***P<.001

TABLE 3

Mean Scores and Standard Deviations of Successful and Unsuccessful
Retrainees on Attitude Scales

SCALES	BEGIN PROGRAM				END ATTITUDE ADJUSTMENT PHASE				END PROGRAM			
	Success N = 44		Failure N = 16		Success N = 44		Failure N = 16		Success		Failure	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
SELF	124.18	25.833	139.13	27.338	131.43	23.295	139.25	27.274	134.70	21.716	145.69	26.32
OTHERS	102.18	11.65	103.13	10.87	100.43	13.25	101.38	14.25	100.36	11.55	104.88	11.67
PATRIOTISM	6.44	1.22	4.99	1.40	6.17	1.42	4.88	1.68	6.34	1.35	5.08	1.54
OBEYING LAW	7.84	.966	6.69	1.41	8.20	1.22	7.24	1.68	7.72	.985	5.93	1.80
AIR FORCE	6.27	2.11	4.13	2.04	6.15	1.85	4.10	1.68	6.02	1.82	3.70	1.96

TABLE 4

Two-way Analysis of Variance of Successful and Unsuccessful Retrainee Scores on

Attitude Scales

SCALES	MS ERROR (df = 174)	MS SUCCESS(A) (df = 1)	F	MS TIME (B) (df = 2)	F	MS A X B (df = 2)	F
SELF	604.04	4342.39	7.19**	1296.29	2.15	193.58	0.32
OTHER	148.15	113.96	.769	43.72	.295	72.88	.491
PATRIOTISM	1.93	62.45	32.36***	.86	.445	.11	.056
OBEYING LAW	1.52	59.55	39.18***	7.36	4.84*	2.20	1.45
AIR FORCE	3.70	165.52	44.74***	1.4	.378	.22	.059

*P<.05

**P<.01

***P<.001

Discussion

The hypothesis that the expressed attitudes of retrainees would become more positive throughout the duration of their rehabilitation program as compared with airmen undergoing normal Air Force technical training was not supported by the data. There were several significant differences between the groups but these differences could not be attributed to the effect of the training itself. There was no significant improvement on retrainee test scores from beginning to end of training on any of the five attitude scales; and predictably the students performed similarly, with only one exception. The student group demonstrated improved attitudes about obeying the law ($p < .01$) but also became more negative in patriotic attitudes ($p < .05$) during the same period of time. The reasons for both changes are rather puzzling considering that nothing in their technical training curricula was designed to directly influence attitudes of any kind and the factors which may have brought about these changes are very unclear. However, the need for additional research to investigate the variables causing such attitude change in a technical course of instruction, where this is not one of the primary objectives, is implicit.

Hypothesis 2 which predicted greater attitude change for the successful retrainees than for the failures could not be supported by the available data. The results failed to indicate an increase in positive attitudes on any of the five scales for either group of men. Although the data yielded little evidence of the effectiveness of the retraining program in making certain attitudes more positive, it does reveal a rather surprising disparity of scores between the successes and failures on the scales measuring attitudes toward patriotism, obeying the law, and the Air Force. The successes had consistently higher scores on all three scales throughout the entire program than did the failures. Moreover, the disparity between the two groups was so great that they could easily be identified and separated simply on the basis of their scores alone, even during their first week in the program. These results have interesting implications for a future prediction study which might try to identify those retrainees who will be most likely to successfully complete the program and be returned to duty.

What does all this mean? Apparently, many men who successfully complete the retraining program do so without appreciably changing their views about certain things they had when they first entered the program. This leads one to believe that attitude change is not as crucial in determining success or failure in the program as perhaps other factors are. Possibly the proper use of these data lies in trying to determine who is likely to benefit most from the program. Thus retrainees with more positive attitudes about patriotism, obeying the law, and the Air Force would be good bets for eventual return to duty and those with more negative scores would be more likely to fail the program.

References

- Aitken, J. R. A study of attitudes and attitudinal change of institutionalized delinquents through group guidance techniques. Unpublished doctoral dissertation, University of Southern Mississippi, 1970.
- Berger, E. The relation between expressed acceptance of self and expressed acceptance of others. *Journal of Abnormal Social Psychology*, 1952, 47, 778-782.
- Bruning, J. L., & Kintz, B. L. *Computational handbook of statistics*. Glenview, Illinois: Scott, Foresman and Company, 1968.
- Brown, B. S. The impact of imprisonment on selected attitudes of recidivists and first offenders. *Journal of Clinical Psychology*, 1970, 26, 435-436.
- Deitz, G. E. The development of self-concept during adolescence and its relationship to adjustment. Unpublished doctoral dissertation, Ohio State University, 1971.
- Gattshall, G. W. Imprisonment's effects upon self-concept and the actualizing process. Unpublished doctoral dissertation, Ball State University, 1970.
- Hamner, W. T. The learning of social controls in juvenile antisocial personalities. Unpublished doctoral dissertation, Vanderbilt University, 1970.
- Kelly, F. J., & Baer, D. J. Jessness inventory and self-concept measures before and after participation in Outward Bound. *Psychological Reports*, 1969, 25, 719-724.
- Thurstone, L. L. Attitude toward patriotism scale. In L. L. Thurstone (Ed.), *The measurement of social attitudes*. University of Chicago Press, 1932.

MILITARY HUMANISM: SOME FURTHER ADVENTURES

IN REHABILITATION

Ogden Brown, Jr.

Air Training Command

The 3415th Special Training Group is a test program to provide increased rehabilitation opportunity for airmen who have been designated unfit or unsuited for further service, and to evaluate the cost, feasibility, and benefits of a centralized program. The program is based upon the principle of conditional suspension of administrative discharge pending successful rehabilitation. A team approach is employed to meet individual needs in an atmosphere of acceptance and understanding. The program is divided into four phases and affords opportunity for attitude development, continuing education, and career development as well as group therapy and individual counseling. Successful rehabilitees have their approved discharges remitted and are returned to duty.

The program of the 3415th Special Training Group (STG) is the result of over ten years of attempts to institute a centralized rehabilitation program for those airmen subject to administrative discharge for cause. Our program is not only to provide increased rehabilitation opportunity for the entire Air Force, but to evaluate the cost, feasibility, and benefits of a centralized program.

The program is based upon the principle of conditional suspension of administrative discharge pending successful rehabilitation. The course of instruction and guiding philosophy of the STG are founded upon a team treatment approach which is designed to meet the needs of the individual in an atmosphere of acceptance and understanding, and yet to allow them to appropriately function within a military environment. As a result, the Air Force hopes to reclaim individuals for further service and, if applicable, to better prepare them for employable roles in the greater civilian community.

Method

Mission

The mission of the STG is to evaluate the feasibility, cost, and benefits associated with a centralized rehabilitation program. Equally important, and more so to the individuals concerned, is to afford the opportunity for a fresh start to those who have been

designated unfit or unsuited for military service. We hope to return these men to productive duty, improved in attitude, conduct, and efficiency.

Organization

The STG is not structured according to conventional Air Force criteria. We have two treatment teams, each headed by a professional psychiatric social worker who is the team chief. The NCOIC of each team is a military training instructor. There are also four other military training instructors, two training technicians, and two psychiatric clinic technicians on each of the two treatment teams. Each of these people is in daily contact with the students assigned to his team. Each team has a maximum of 20 students assigned at any given time, and replacement is on a one-for-one basis. We also have an Evaluation and Treatment Branch which is headed by a psychiatrist and includes a clinical psychologist as well as another psychiatric clinic technician.

Eligibility

Only enlisted members of the Air Force, approved for discharge under AFM 39-12 for unsuitability, misconduct, or unfitness, are eligible for probation and rehabilitation. Each must be granted probation and rehabilitation by his discharge authority and, in addition, must volunteer to come to the centralized program. Such volunteers are nominated, and selection is accomplished from these nominations submitted by field units.

Program

The program of the STG is presently divided into four phases. We began with three phases, but added a fourth phase which we termed an honor phase wherein the student is not restricted as to hours nor is he subject to any bed check.

Phase One is our orientation period and is conducted during the student's first week. In-processing, testing, and interviews are conducted during this phase. Students are restricted to the immediate Group area.

Phase Two, lasting four weeks, is concerned with helping the student to better understand himself, his motives, and his actions. The purpose of this phase is to effect a change of attitude in the student which will allow him to return to productive duty and, hopefully, to become a better citizen. We approach these aims through classroom instruction, group counseling sessions, and individual counseling as well. While undergoing this attitude development and adjustment phase, the student is restricted to the confines of the base.

Phase Three, having no specific temporal duration, is more personalized and is dependent upon the needs of the individual himself. Those found low in academic achievement are enrolled in remedial classes in mathematics, English, social studies, or reading improvement in order to raise their educational achievement level. We also offer the high school General Education Development program, Extension Course Institute, and United States Armed Forces Institute programs. Career Development is also individualized: the student may enter Technical School, he may be placed on a job in his career field, or he may be cross-trained into a new field. The type and length of training are flexible, but each student will attain at least a three level AFSC before graduation. We use this phase to afford the student an opportunity to employ his changed life style in a realistic environment. We watch him closely and obtain frequent reports on his behavior and progress. Students have privileges during this phase which allow them to leave the base, but they must return for bed check at certain prescribed hours.

Phase Four, the honor phase, differs from Phase Three only in the fact that the students are given unrestricted hours. They enjoy off-base privileges, and the only constraint is they must report to their duty assignments when scheduled. In this manner we not only reward desired behavior, but we can again observe the airman in as real life a situation as is possible. We would rather have any unacceptable behavior occur while the man is still undergoing rehabilitation, not after he has been prematurely returned to duty. Phase Four allows us to better assess such possibilities, and to deal with such behavior in a less demanding environment.

Team Treatment

Each student is assigned to one of the two treatment teams. He is assigned a team member as his individual counselor, one to whom he can go for help in every phase of the program. The team is our basic unit; the student interacts with the team from the day he arrives until the day he leaves. Each student also meets with the entire team every two weeks. At these meetings, information is gathered, advice is given, and decisions are reached concerning student progress.

The primary purpose of team treatment is to help the student better understand himself, and to recognize what it is within him that led to administrative discharge action. Acceptance and understanding on the part of every member of the team are very important in this process. Further, the student must realize that people are there to help him, that he can change for the better, and that there is a place for him in the Air Force. In an atmosphere of interest and genuine concern, then, the student is helped to look inward, to gain insight, and to learn a new, acceptable mode of behavior for the future.

We hold that both dynamic and learning models are relevant in our team treatment approach, which is partially oriented towards increased intra- and interpersonal effectiveness. Our methodology is holistic and is basically addressed toward the extinction of inappropriate behavior and the reinforcement of more acceptable behavioral responses.

Results

Student Disposition

When the treatment team has decided that a man is ready to return to duty, a Team Adjustment Board is convened. If a favorable decision is reached by this board, the recommendation is made for a Final Group Adjustment Board. If the results of this action again are favorable, the board then recommends to the Group Commander that the student be returned to duty, subject to any constraints which might apply for rehabilitative purposes. The Group Commander reviews and evaluates the entire case and then presents his recommendation to the final approving authority, the Lowry Technical Training Center Commander. If he concurs, the approved administrative discharge is remitted and the student is returned to full duty. The alternatives are, of course, to continue rehabilitation or to return the man for execution of his discharge.

If a team feels that an airman is not proceeding satisfactorily, or if he is guilty of repeated misconduct while undergoing rehabilitation, the team chief is authorized to hold a Team Disciplinary Board. A Group Disciplinary Board may then follow in recommending action to the Group Commander. While undergoing rehabilitation, a student is subject to courts-martial, actions under the UCMJ, and other disciplinary measures. If it is decided that a student does not enjoy the necessary potential for rehabilitation, the Group Commander recommends to the Center Commander that the student be returned to his parent unit for execution of his suspended discharge. If appropriate, however, new AFM 39-12 or courts-martial actions may be taken.

In summary, then, the Center Commander has the authority to remit the administrative discharge if the student successfully completes his rehabilitation program. He also has the prerogative, if appropriate, to terminate the student's TDY and return him to his original organization with the recommendation that his suspended discharge be executed.

Cross-Training

Students whose problems are related to their job or to their career field, or who cannot perform duty in their career field because of human reliability or personnel reliability, are allowed to cross-train into other career fields. This training is done under the supervision of the training technicians on each team. It is usually

accomplished through on-the-job training at various job outlets on Lowry AFB or nearby Buckley Air National Guard Base. We attempt to provide the opportunity for each airman to reach the three level skill in his new career field.

Discussion and Summary

Results

The first airman to arrive at the STG signed in on 2 May 1971. As of 1 April 1972, 102 airmen have entered the program. Of this number, 65 have been released from training; 46 were returned to full duty, and 19 were unsuccessful in their rehabilitation training. Twelve of the unsuccessful students maintained their continued frequent involvement with authorities; one was found to be continuing his illegal use of drugs; two were convicted by courts-martial; one was referred to medical channels; one had his TDY terminated because he went AWOL; and the last, an alcoholic, could not kick his habit.

We have begun data analysis in order to make valid recommendations to higher authorities and to evaluate the feasibility, cost, and benefits associated with a centralized rehabilitation concept. Our data are all preliminary at this point, but certain trends and indications are emerging. Therefore, some speculation would seem to be appropriate.

Among the 102 students, 21 enjoyed honorable discharges and 21 had undesirable discharges (UD). One might infer that those with honorable discharges desired to remain in the Air Force and that those with the UD's desired to have that stigma removed. Interestingly, only three of the 19 unsuccessful students had UD's to be executed.

The most common reason for discharge among our students is the category entitled character and behavior disorder (n = 22). Close behind are the categories of frequent involvement with authority (n = 20) and misconduct because of civil court disposition (n = 17). We have had 16 airmen referred for apathy or defective attitude, and 12 have received their discharges for drug abuse.

When the type of discharge is considered in conjunction with the reason for discharge, several interesting facts are revealed. For example, no honorable discharge was awarded to either a drug abuser or to one who has been frequently involved with authority. One might speculate, then, that the latter's behavior is not condoned as socially acceptable, while the airman who has been continually rejecting authority is likely to be the one who is the classic "troublemaker", at least in the eyes of the commander who recommended that he be discharged. Further, no UD's were awarded for character and behavior disorders, inaptitude, or for apathy and defective attitude. The inference here

is that these categories are perhaps beyond the immediate control of the individual himself and thus not really his fault. To reinforce this inference, AFM 39-12 specifies that an honorable or general discharge will be issued for these conditions of unsuitability. Discharges for unfitness, however, may include all three types.

Our students have ranged in grade from Staff Sergeant down to Airman Basic. These grades have been normally distributed and correlate with time in service with the exception of Airman Basic. About one-third of our students have been, at some time or another, reduced to the grade of Airman Basic. None of them came from basic training; all at one time or another had held a higher grade. It is informative to note that half of the Airmen Basic have their discharges for frequent involvement. This fact seems to parallel our earlier inference regarding the fact that this is the airman who is the "troublemaker" in the eyes of his commander.

One further observation which appears to be useful is that, in the category of frequent involvement with authority, we have returned more students for execution of their discharges than we have returned to duty. The sample is, of course, much too small for any valid generalization or inference to be drawn. It does appear, however, that the prognosis for success in the case of frequent involvement is rather poor. On the other hand, the prognosis for those convicted by a civil court and for those categorized as character and behavior disorders appears to be quite favorable.

Follow-Up and Evaluation

We follow the successful graduate by sending out a rating scale and questionnaire to both his immediate commander and the first-line supervisor. We request these at three, six, nine, and twelve months after successful completion of the program. In this fashion we are in a position to ascertain whether the overt rehabilitation was only temporary or whether we indeed achieved permanent results. We have been extremely pleased to learn that the average rating given by the immediate supervisor has been 7.2 on a 9 point scale at the six month follow-up point. We have also found that the ratings given by the commanders concerned have averaged 7.8 on a 9 point scale at the six month follow-up point. These average ratings suggest to us that perhaps the outcomes of our rehabilitation training might indeed be more than just a transitory behavioral modification.

Conclusion

The 3415th Special Training Group is, of course, a test program. We know at this time that we will be successful in a number of cases. The real question is not whether rehabilitation can be done successfully, but whether such a program is of real value to the Air Force.

That it is feasible we have no doubt. The question of whether such a program is cost effective and whether the benefits accruing to the individual, the Air Force, and the larger society are worth the cost remains to be answered. Based upon our preliminary results at this point, this appears to be the case. We are encouraged by the results we have achieved to date, and hope that the program indeed will be of real value to all concerned.

2 - *Human Factors (OAR Sponsored Research)*

CHAIRMAN: Major Lawrence F. Sharp

PREDICTIVE VALIDITY OF GROUND-BASED FLIGHT CHECKS

Jefferson M. Koonce

Aviation Research Laboratory, University of Illinois

The first of three phases of work on the predictive validity of ground-based flight checks sponsored by the AFOSR is reported. Phase I is concerned with the performance of private pilots in light single-engine aircraft. An overview of Phase II and III is included.

The evaluation of a pilot's performance is typically done by means of inflight check rides requiring the expense of the aircraft and exposure to certain additional hazards. Some commercial airlines, for whom cost is an essential factor, have been using ground-based simulators for training with a high degree of success. The purpose of this study is to determine the predictive validity of ground-based flight checks upon subsequent performance in aircraft.

Many attempts have been made to develop objective-type scoring booklets for evaluating pilot proficiency in flight (Gordon, 1949; Smith, Flexman, & Houston, 1952). The purpose in developing these scoring methods was to measure more accurately the subject's performance by restricting the amount of subjectivity involved in the scoring decisions. In this study, the Illinois Private Pilot Performance Scale (IPPPS) was used to minimize observer-observer unreliability. On-going research using the IPPPS has been yielding observer-observer correlations in the high .80s for inflight maneuvers.

Because the ability to predict subsequent inflight performance from ground-based simulated check rides is influenced by the pilot's consistency of performance and the change in hardware from simulator to aircraft, this study also looked at three additional conditions: simulator-simulator, aircraft-simulator, and aircraft-aircraft.

Method

Equipment

The simulator used in this study was a Link GAT-1 (GAT) modified so that two observers could evaluate a subject's performance at the same time. Cessna 172s and Piper Cherokees, depending upon the subject's preference, were used for evaluation of the subject's inflight performance.

The observers making the evaluations of the subject's performance were nine flight instructors from the University of Illinois'

Aviation Institute. For inflight data collection, one observer rode in the front right-hand seat of the aircraft as a safety pilot while a second observer sat in the right rear seat, behind the safety pilot with a good view of the instrument panel.

Subjects

A total of 166 volunteer private pilots with varying experience from forty hours to 13,000 hours flight time were used as subjects. The GAT-Aircraft condition had 30 Ss, the GAT-GAT condition had 51 Ss, the Aircraft-GAT condition had 35 Ss, and the Aircraft-Aircraft condition had 50 Ss assigned. The differences in the numbers of subjects per condition were primarily due to weather conditions and hardware availability.

• 3.

Design

A subject pilot was evaluated on each of two successive days, those rides corresponding in order to the name of the condition to which he was assigned. On Day 1 the subject pilot was briefed and then flew all of the maneuvers in the IPPPS (1st Attempt) followed by a second performance of all the maneuvers (2nd Attempt). During both the first and second attempts at the maneuvers the pilot's performance was evaluated by two observers (Observers 1 and 2 on Day 1). On the following day the subject performed all of the maneuvers again two times, first and second attempts of Day 2, and again was evaluated by two observers (Observers 2 and 3 on Day 2). On Day 2 one of the observers was new, having not flown with the subject pilot before; and the other observer was one of the two who flew with the subject pilot on the previous day.

Thus, a subject of the GAT-Aircraft condition would fly the GAT on Day 1 while being evaluated by Observers 1 and 2 through two performances of the maneuvers in the IPPPS. On Day 2 the subject would perform all of the maneuvers of the IPPPS twice in an aircraft while being evaluated by Observers 2 and 3.

IPPPS

The IPPPS is a booklet of eleven maneuvers to be evaluated each having between four and six criteria measures, giving a total of 47 criterion measures for one observer scoring one attempt. With the subject performing each maneuver twice each day and being evaluated each time by two observers a total of 188 criterion measures were taken on each day of the Aircraft-Aircraft condition. For the conditions in which the GAT was used there were fewer criterion measures because maneuvers such as a 720 degree turn around a point, and takeoffs and landings could not be adequately scored or performed.

Results and Discussion

The scores for the individual criterion measures were coded on computer cards and converted to standard scores by means of a computer program. Individual criterion measures for a particular maneuver were summed to yield a score for the maneuver. All of the maneuver scores for an attempt were summed to give a composite score for one observer on one attempt. Similarly, all of the composite scores for the observers and attempts were combined to give a Day 1 composite score to be compared with the Day 2 composite score for the subjects.

Because of a lack of detailed examination of the earlier use of the IPPPS the individual criterion measures and maneuvers were not weighted. Pearson product-moment correlations were used to determine the relationships among performances scored under various conditions, attempts, and observers.

The correlations of the subject's composite score for Day 1 with his composite score for Day 2 were as follows: Aircraft-Aircraft = 0.81, Aircraft-GAT = 0.51, GAT-Aircraft = 0.56, and GAT-GAT = 0.84.

Table 1 shows the correlations between observers for the four conditions as quite high. Note that the correlations of performance between Day 1 and Day 2 for the GAT-GAT and the Aircraft-Aircraft groups are in the .70s and .80s while those of the GAT-Aircraft and Aircraft-GAT conditions are in the .40s and .50s.

In both the overall composite score correlations and those given in Table 1 the correlations are lower for those conditions for which there was a hardware change, GAT-Aircraft and Aircraft-GAT. The correlations between the two observers scoring the same attempts was very high (.85 to .94) which indicates low observer-observer unreliability using the IPPPS.

In the analysis, the various flight maneuvers evaluated were divided into two types: contact--those being performed with reference to the visual world outside the aircraft, and instrument--those performed solely by reference to the flight instruments inside the aircraft. Table 2 indicates that there is much higher agreement in the scoring of instrument maneuvers than the contact maneuvers. This was as expected.

The analysis of the data is continuing and an item analysis of the individual maneuvers and criterion measures will be made to determine the extent to which they contribute to the predictability of subsequent performance.

TABLE 1

Correlations for Observers 1 and 2 on Day 1
and Observers 2 and 3 on Day 2
for the Four Conditions

Aircraft - Aircraft					GAT - Aircraft				
1st Day		2nd Day			1st Day		2nd Day		
Obs	1	2	2	3	Obs	1	2	2	3
1st Day	1	.91	.74	.78	1st Day	1	.85	.52	.48
	2		.70	.77		2		.52	.58
2nd Day	2			.87	2nd Day	2			.90
	3					3			

GAT - GAT					Aircraft - GAT				
1st Day		2nd Day			1st Day		2nd Day		
Obs	1	2	2	3	Obs	1	2	2	3
1st Day	1	.94	.84	.84	1st Day	1	.88	.51	.50
	2		.76	.80		2		.49	.43
2nd Day	2			.93	2nd Day	2			.88
	3					3			

Phases II and III

The second phase of this effort is concerned with predicting the inflight performance of instrument rated pilots flying a brief IFR flight plan that includes three approaches. The scoring booklet has been developed and the experimental design is similar to that used in Phase I. Two contact maneuvers are included in Phase II, takeoff and

TABLE 2

Correlations of the Instrument (I) and Contact (C) Maneuvers
for Day 1 of the Aircraft-Aircraft Condition

			1st Attempt				2nd Attempt			
			Obs 1		Obs 2		Obs 1		Obs 2	
			I	C	I	C	I	C	I	C
1st Attempt	Obs 1	I		.22	.91	.41	.79	.33	.80	.23
		C			.30	.49	.18	.26	.27	.18
	Obs 2	I				.38	.78	.29	.83	.24
		C					.44	.39	.51	.62
2nd Attempt	Obs 2	I					.28	.91	.26	
		C						.27	.44	
	Obs 3	I								.31
		C								

landing, and the criterion measures for these have been modified slightly. The aircraft and simulators for Phase II are the same types used in Phase I.

Phase III will determine the predictive power of ground-based check rides for experienced pilots within an operational framework. Twin engine aircraft and the GAT-2 simulator will be used in the third phase.

References

- Gordon, T. A. *The Development of a standard flight-check for the airline transport rating based on the critical requirements of the airline pilot's job.* Washington, D. C.: CAA Division of Research, Report No. 85, 1949.
- Smith, J. F., Flexman, R. E., & Houston, R. C. *Development of an objective method of recording flight performance.* Lackland AFB, Texas: Human Resources Research Center, HRRC Technical Report No. 52-15, 1952.

MOTION CUES AS A FACTOR IN SIMULATOR AND AIRBORNE
EVALUATION OF FLIGHT DIRECTOR DISPLAYS

Robert S. Jacobs

Aviation Research Laboratory, University of Illinois

The results of research conducted in ground-based flight simulators must be interpreted with care in view of the effect upon such findings of differences between the simulator and simulated internal environments. This paper describes one study which produced significantly different results as a function of motion simulation fidelity. Implications of this finding are discussed.

Ground-based flight simulators have been effectively applied to three basic classes of aviation activity; training, performance evaluation, and research. Transfer studies have been conducted which demonstrate that the simulator is an efficient and economical tool for pilot training (Povenmire & Roscoe, 1971). Unfortunately, however, there have been few studies which investigate the broad issue of the validity and limitations of simulators of various degrees of fidelity as environments for evaluation and research. Where experimental evidence has shown that the utility of a training simulator is relatively independent of simulation fidelity (Briggs, Fitts, & Bahrick, 1957, 1958), the value of study results obtained in flight simulators may hinge more closely upon the accuracy of reproduction of simulated environment. As in the case of an aircraft in flight, this environment must give rise to an exceedingly complex and varied assortment of informational cues.

It is worthwhile to consider the nature of the system involved when experimentally evaluating flight displays in a simulator. It is a man/machine ensemble which is assumed to behave in the same manner as would a system composed of this same man, and an aircraft in flight. This man, however, is a highly trained information processor. He has been conditioned to act upon certain information received from his environment through particular channels in a very particular fashion. In flight, an aircraft responds to pilot actions in accordance with physical laws that constrain the response dynamics. In general, simulators are built to mimic aircraft responses artificially, using computers to calculate the trajectory of response, and various displays (not necessarily visual displays) to communicate these responses to the simulator pilot. What are the results if the simulator doesn't supply this feedback information through the same channels and in the same scale as the pilot is accustomed to in the air? In such instances, the informational environment of the simulator will

differ from the informational environment of the simulated aircraft. Because the pilot is trained to respond on the basis of this environment, it is reasonable to expect that the pilot's responses in the simulator will differ from his responses in the air. These differences may be subtle, they may be insignificant with respect to research objectives, but it seems clear that to some degree they will be reflected in the data collected. This is not to say that such experimentation cannot be generalized to the airborne situation, rather, that considerable caution must be exercised in doing so.

The uncertainty introduced into research findings by imperfect fidelity of cue environment in flight simulators is particularly acute with respect to performance dependent upon the general class of cues associated with the kinesthetic and vestibular senses. Ground-based flight simulators are incapable of generating a realistic reproduction of the linear and angular accelerations which may be experienced in flight. While this discrepancy has been acknowledged in the literature covering simulator research, there have been few attempts to examine parametrically the effect of varying the fidelity of motion simulation in order to understand its effect. That such studies have not been made incidental to applied investigations on simulators with very high levels of fidelity is understandable, as the investigator tends to use the full capability at hand in the implicit assumption that such a course produces results more akin to flight. This assumption may not be valid in all instances. Since the simulator is restricted in terms of its response amplitude and frequency capabilities; and even in terms of its degrees of freedom, the motion cues it generates may in subtle ways be false or misleading. These wrong cues may be, in fact, more distorting of experimental results than no motion at all. Further, in many task situations, motion cues do not materially add to the richness of the informational environment. In such situations, motion cues become competing noise rather than signals, and may add unwanted variation to results due to diversion of subject attention necessary to overcome irrelevant sensations.

One study which considers the effect of motion fidelity upon simulator research results has been reported by Matheny, Dougherty, and Willis (1963). Two attitude indicator displays, one with a moving aircraft symbol, and the other with a moving horizon symbol, were compared under two conditions of simulator motion. The task involved for the subjects was simply interpretation of the display indication for various states of pitch and roll orientation. The results indicated that there was a statistically significant difference in the order of merit results obtained between the two conditions. Without motion, naive subjects were more accurate in their interpretation of the moving airplane indicator than in reading the moving horizon. When motion cues were added, the difference disappeared, but response latencies were significantly reduced for both conditions. In their

conclusions, the authors state, "...It is evident that motion is an extremely relevant variable in the evaluation of displays in situations in which motion cues are present." and that, "...the degree to which it (the simulator) duplicates the angular motions of the vehicle being simulated is important. Lack of motion cues may lead to erroneous conclusions as to the suitability of displays for systems in which motion cues are present."

These results sound a cautionary bell for those interested in comparative evaluation of attitude displays. One must determine the appropriate environment for such a study as incorrect conclusions can result from improper selection. The Matheny, et al. (1963) study involved very simple motion cues and a simple interpretation task. A task situation involving closed loop feedback through the display with complex motion cue structure could be expected to be even more critically influenced by the fidelity of the motion simulation. If principles generalizable to airborne application are being sought, it would seem that with such complex interactive tasks, the risk of arriving at an incorrect conclusion in a simulator might be even greater.

Method

Under contract to the Air Force Office of Scientific Research and the Office of Naval Research, the Aviation Research Laboratory of the University of Illinois has been conducting a multiple phase comparative evaluation of a number of common and experimental motion relationships for symbolic flight director/attitude indicator displays. Mindful of the probable interactive effect of the motion cue structure of the experimental environment, it was proposed that the study be repeated under three conditions of motion fidelity. The first of these was conducted in the Laboratory's Link GAT-2 simulator with the simulator's motion system in normal operation. This first effort was then duplicated in the GAT-2 with the motion system off. A third environment under current investigation is actual flight. The vehicle for provision of this environment is the ARL's "flying laboratory," a Beech C-45H. While a systematic comparison of the results for each of these environments must await completion of the inflight study, some interesting conclusions may be drawn by analysis of results from the first two conditions.

Experimental Displays

The eight experimental flight director display configurations studied represent the various combinations of four basic modes of attitude presentation and two basic modes of command guidance presentation, compensatory and pursuit. These four basic attitude presentation modes were; conventional moving horizon (inside-out), moving airplane (outside-in), kinalog (time-lagged frequency separation), and a hybrid

frequency-separated presentation using aileron position to quicken the indication of bank attitude changes. All eight display configurations employed the same three symbols on the CRT presentation, and the same signals and scale factors combined in various ways to drive the three symbols. The symbols were (a) a segmented line representing the horizon, (b) a two-line symbol representing a cross-sectional view of the airplane from the rear, and (c) a dot presenting horizontal and vertical steering commands. The eight displays and the signals that drive the respective symbols are illustrated in Figure 1.

3

Experimental Design

Two tasks were given each subject, a practice task intended to familiarize him with the display dynamics and to allow him to reach a stable level of performance, and an evaluation task. The practice task consisted of steering out a series of step commands in the display horizontal dimension while tracking out Gaussian noise in the vertical dimension. The evaluation task required the subject to track out continuous Gaussian perturbations in both the horizontal and vertical display dimensions. In both tasks, the Gaussian noise had a cutoff frequency of 0.05 Hz.

The procedure for the practice task was as follows. The displays were introduced prior to practice with each one by a one-minute demonstration with verbal explanation by the experimenter. The subject was then permitted a total of 9.6 minutes of practice divided into three equal periods separated by short breaks. Subjects were shown 4 display conditions on each of two days on the practice task so that at the end of the second session they had seen all 8 conditions. The order of presentation in the practice task was counterbalanced to minimize transfer effects.

The evaluation task was performed approximately 24 hours after the second session of the practice task. It consisted of one three-minute trial on each of the 8 display conditions with one-minute rest periods between successive trials with the exception of a five-minute break between the fourth and fifth trials. The order of presentation of the conditions was again counterbalanced to minimize transfer effects.

Subjects

Sixteen private pilots with experience ranging from 40 to 150 hours of flight time were used as subjects. The motion variable was between subjects, so that 8 subjects were tested on all displays with simulator motion, and the remaining 8 were tested on all displays without simulator motion. All subjects had been trained by and had earned their pilot's licenses at the University of Illinois' Institute of Aviation.

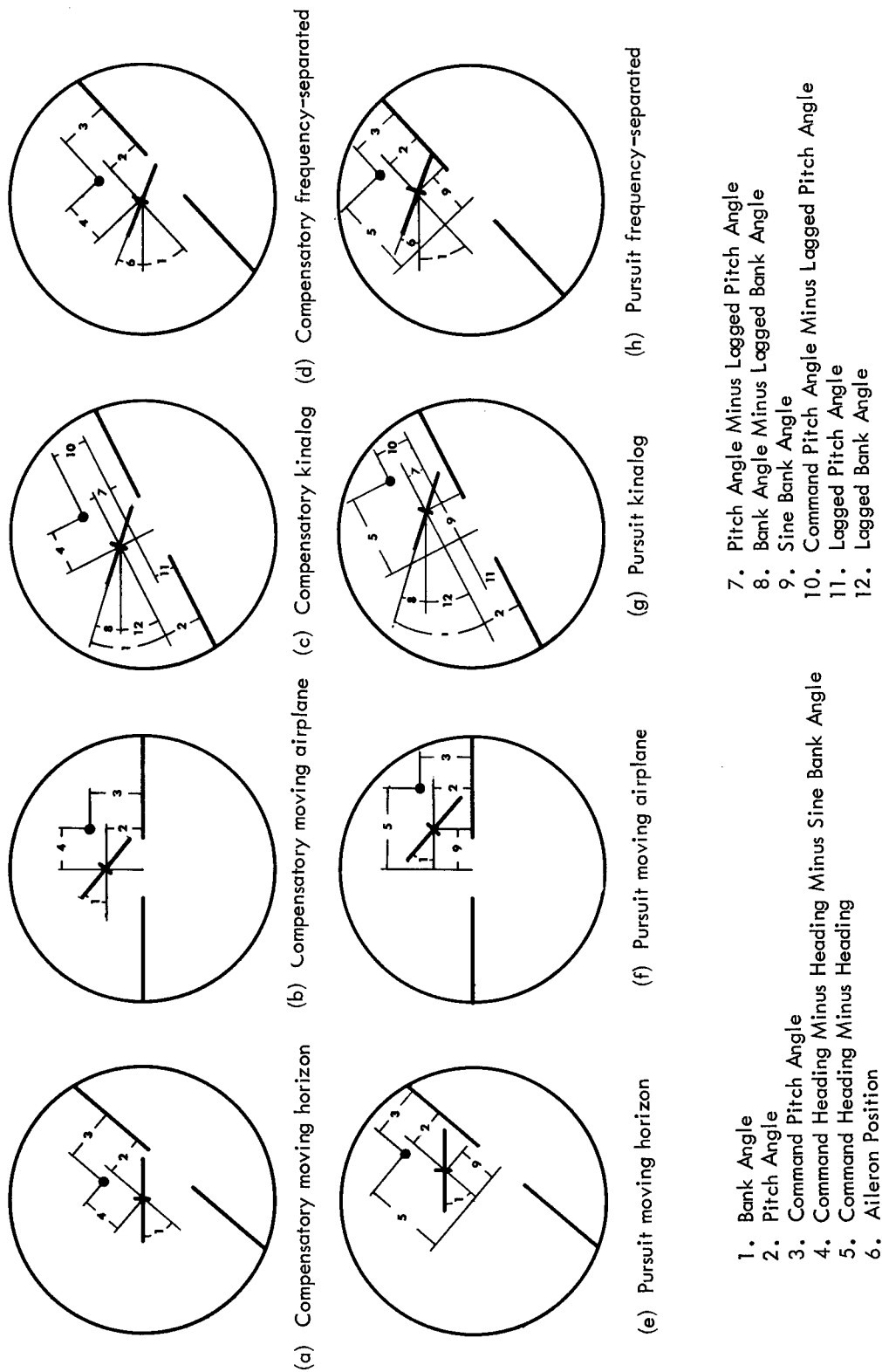


Figure 1. Experimental display configurations.

Results

Horizontal steering performance in the evaluation task for each of the 8 display conditions, with and without simulator motion is presented in Table 1. The values shown are the means of the logarithms of the roots of the integrated squared errors in azimuth.

TABLE 1

Mean Log RMS Azimuth Steering Error for Independent Groups of Pilots
Flying Each of Eight Displays with Motion System
ON and OFF

Display	Log RMS Error	
	Without Motion	With Motion
Compensatory Moving Horizon	.818	.771
Compensatory Moving Airplane	.868	.761
Compensatory Kinalog	.827	.747
Compensatory Frequency Separated	.829	.751
Pursuit Moving Horizon	.840	.758
Pursuit Moving Airplane	.753	.709
Pursuit Kinalog	.748	.751
Pursuit Frequency Separated	.776	.754

Analysis of variances showed the following results to be significant.

1. For all displays, pilot performances in azimuth steering were better with simulator motion than without, and there were performance differences among displays under both conditions.
2. There was an interaction between command steering presentation and attitude presentation.
3. There was an interaction among command steering presentation, attitude presentation, and the presence or absence of simulator motion.

TABLE 2

Analysis of Variance Summary of Log RMS Horizontal Tracking Error
Comparing Display Formats in the
Presence or Absence of Simulator Motion

Source	df	MS	F
Between Subjects			
Motion (M)	1	0.1060	6.07*
Subjects (S/M)	14	0.0175	
Within Subjects			
Displays (D)	7	0.0124	4.02**
D X M	7	0.0052	1.69
D X S/M	98	0.0031	

* $p < .05$

** $p < .001$

The first finding serves to confirm the results of studies such as Matheney, et al., that have shown simulator motion to facilitate manual control performance. The second result confirms the findings that manual control is disproportionately superior with the pursuit moving airplane display. Of greater interest however is the third finding, because it demonstrates that motion cues can indeed affect the outcome of display or control evaluations, and it offers a basis for reconciling apparently contradictory results from previous studies of pursuit and compensatory tracking.

The precise role of motion cues in influencing human behavior in flight simulation task situations is not well understood, just as the role these cues play in the inflight piloting process is not completely clear. That they influence this behavior is clear as we have seen that manipulation of these cues changes performance levels observed in simulated piloting tasks. When conducting comparative evaluations of the type described here, the usual technique is to exercise the system in an operational function and to examine the obtained performance as the experimental variable (in this case, type of display) is varied.

TABLE 3

Analysis of Variance Summary of Log RMS Horizontal Tracking
Error Comparing Attitude and Command Mode Presentations
in the Presence or Absence of Simulator Motion

Source	df		F
Between Subjects			
Motion (M)	1	0.0859	7.56*
Subjects (S/M)	14	0.0114	
Within Subjects			
Attitude Format (A)	2	0.0076	2.46
A X M	2	0.0029	.94
A X S/M	28	0.0031	
Command Format (C)	1	0.0360	6.41*
C X M	1	0.0086	1.52
C X S/M	14	0.0056	
A X C	2	0.0152	7.32**
A X C X M	2	0.0078	3.76*
A X C X S/M	28	0.0021	

* $p < .05$

** $p < .01$

There is risk in generalizing the findings of these studies to operational systems if the effects of the differences between the operational environment and the research environment are not both understood and compensated for in the application of these findings.

References

- Briggs, G. E., Bahrick, H. P., & Fitts, P. M. The influence of force and amplitude cues on learning and performance in a complex tracking task. *Journal of Experimental Psychology*, 1957, 54, 262-268.
- Briggs, G. E., Fitts, P. M., & Bahrick, H. P. Transfer effects from a single to a double integral tracking system. *Journal of Experimental Psychology*, 1958, 55, 135-142.
- Johnson, S. L., & Roscoe, S. N. *Frequency separated flight displays*. University of Illinois, Aviation Research Laboratory Report ONR-70-2/(ARL-70-12), October 1970.
- Matheny, W. G., Dougherty, D. J., & Willis, J. M. Relative motion of elements in instrument displays. *Aerospace Medicine*, 1963, 34, 1041-1046.
- Povenmire, H. K., & Roscoe, S. N. An evaluation of ground-based flight trainers in routine primary flight training. *Human Factors*, 1971, 13(2), 109-116.

SPATIAL AND TEMPORAL ASPECTS OF PERCEPTION

WITH VISUALLY TIME-COMPRESSED DISPLAYS

Lawrence A. Scanlan

Aviation Research Laboratory, University of Illinois

Previous research has demonstrated the effectiveness of the coherent motion cues provided by a visually time-compressed radar display. The results of these studies suggest that further improvement may be possible by combining spatial cues with the temporal motion cues. A preliminary study was conducted to verify the procedure and apparatus to be employed in a test of combined spatial and temporal cues. Results of this study identified two aspects of the experimental task which need to be investigated before an adequate test of combined spatial and temporal cues can be made.

Recent studies aimed at improving an operator's ability to detect radar targets in the presence of noise and clutter have demonstrated the effectiveness of a display presentation technique best described as visual time compression (Scanlan, 1971; Scanlan, Roscoe, & Williges, 1971). The detection-enhancing effect depends upon a Gestalt quality of man's visual perceptual system called the Phi phenomenon by Wertheimer (1912).

The task of detecting a target on a ground-based radar consists primarily of discriminating coherent target motion from both the non-coherent, randomly appearing noise and the clutter of stationary returns from the surrounding terrain. On an airborne radar display, the clutter also has coherent motion but of a rate different from that of airborne targets. A time-compressed display accentuates the coherent motion of targets relative to the random noise and slowly moving clutter to yield improved detection performance.

A time-compressed display is obtained by storing several past image frames and playing them back in the order in which they were collected but at a faster rate. If these frames are repeatedly played back at an appropriately fast rate, returns from a moving target appear as a rapidly moving dot traversing the display. The dot, first evident in the oldest frame, moves across the display until it appears in the most recent frame. The dot motion then starts over, appearing on the oldest frame again, and retraces its path.

While this sequence is happening rapidly, the display is updated with new information gathered in real-time. Each new frame replaces the oldest frame so that only the desired number of preceding frames

is stored for display. The overall effect is that of a repetitive moving dot sequence that slowly advances across the display. The rapid motion of the dot adds to the conspicuousness of the target, while the slower motion of the coherent dot sequence keeps the target position current.

Scanlan, Roscoe, and Williges (1971) investigated visual time-compression using a simulated radar display that presented targets, noise, and clutter as bright dots on a cathode-ray tube. The number of stored image frames and the rate at which they were played back were manipulated along with several other variables.

Three factors, number of frames stored, time-compression ratio, and noise level, were found to have pronounced effects on the time required to detect a target. A fourth factor, clutter, caused only a slight change in time to detect. An analysis of variance of these data indicated that all of these effects would be expected to occur by chance less than once in a hundred replications of the experiment ($p < .01$).

Figure 1 graphically presents the effects of time-compression ratio as a function of the number of stored frames. In this figure a time-compression ratio of unity corresponds to a standard radar display while a time-compression ratio of infinity corresponds to a condition in which all of the stored frames are shown virtually at one time. In the latter condition the target motion cue, present at all intermediate time-compression ratios, is absent, and a spatial pattern is the effective detection cue.

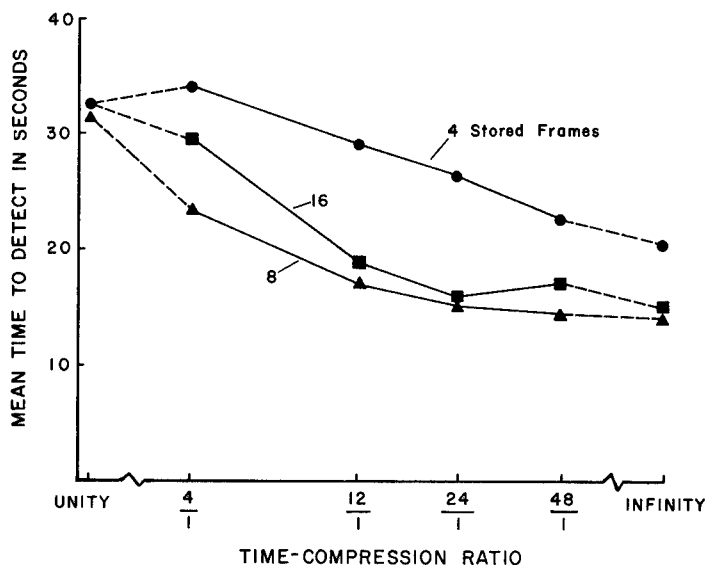


Fig. 1. Effect of time-compression ratio and number of stored frames on time required to detect a target based on the means of six trials by each of 18 subjects.

Although several aspects of these data have implications for further research, the remainder of this paper will concentrate on only one. An examination of Figure 1 indicates that if enough image frames are stored, 8 or 16, and the time-compression ratio is high enough, greater than 12 to 1, a dramatic increase in performance over a standard radar presentation is obtained. A similar reduction in the number of missed and falsely detected targets is simultaneously obtained.

These data are clearly in accordance with the hypothesis that coherent motion is a powerful and effective cue in a detection task. What is not so readily apparent is the high performance obtained when the time-compression ratio equaled infinity. In this condition all frames are merged together and presented simultaneously, thus removing the motion cues and replacing them with spatial pattern cues. This would seem to argue for an equivalence of spatial and temporal cues in perception. Such a conclusion does not, however, accord well with existing theories of visual perception, and a search for alternative explanations is necessary. One possible alternative can be found in the particular method used to introduce the target.

The subject was initially shown a display with noise and clutter but no target. The target was introduced with a variable delay not exceeding 20 seconds in accordance with a list of random numbers. Such a method is reasonable from a real-world standpoint where it is not known when or where the target will appear. However, with time-compression ratios other than unity and infinity there is a period of time following the introduction of the target during which returns are present only a fraction of the time. For example, with a time-compressed display with 8-frame storage, when the first target return is received, it will be stored as part of the most recent image frame. As the stored frames are played back one at a time, all but the most recent will contain only noise and clutter, and only the most recent will present the target. In other words, the target will be displayed with a duty factor of only one-eighth. When the second return is received, two frames will contain target information, and the observer can process target returns one-quarter of the time. Continuing this process with four returns, the observer may see the target no more than half the time, and only when eight returns have accumulated can he observe the target all of the time. With 8-frame storage, this means that more than 13 seconds are required before the display can be fully effective. Contrast this with the infinite time-compression ratio in which all available information about the target is presented all of the time. If the target is present only on the four most recent frames, it will appear as four dots that can be observed all of the time.

Because of the target duty factor effect, the Scanlan, Roscoe, & Williges (1971) experiment does not provide any evidence as to the relative contribution of spatial and temporal cues in perception. However, for the applied problems of optimizing a radar display the Scanlan et al. data are germane and suggest that the optimum display may be one that combines spatial and temporal cues.

At least two techniques are available for obtaining a hybrid that combines the apparent motion cue and the nearly full-time presentation of all available target information. One method would be to modify the playback technique to produce a growing trail. The oldest frame would be displayed, and then the next oldest would be added, leaving the original frame displayed rather than replacing it. Each frame would be added until all stored frames were displayed. Then all frames would be removed, and the sequence would begin again. This would produce a rapidly lengthening line of dots that would have both apparent motion and spatial pattern and would present all available information about the target only slightly less than full time.

An alternative method would be to keep the normal time-compressed presentation but at the end of each playback sequence pause and show all of the stored frames. The display would then alternate between an intermediate and infinite time-compression ratio. The playback rate or time-compression ratio and the length of the pause could be varied independently to obtain an optimum combination.

Preliminary Experiment

A preliminary experiment was performed to verify the procedure and variable levels to be used in a larger study which will test the hypothesis that a combination of temporal and spatial cues will improve detection performance. Because this was also the first time the computer-controlled videotape system was used, this preliminary study served as a test of that system.

No statistical analysis was attempted because only three subjects were tested. Each of the subjects was given four trials in each of six treatment combinations. The treatments were the factorial combinations of the three modes and two playback rates (time-compression ratios). The noise level was held constant at 48 per frame and the number of stored frames at 8.

Method

Apparatus

Subjects were shown video-tape-recorded TV presentations of a simulated radar display and asked to discriminate a target from the

random noise also present on the display. The videotapes were made by recording the output of a modified Hughes Aircraft Company digital scan converter capable of generating the time-compressed display and the two variations discussed above. A computer-controlled videotape system was used to play the tapes for the subjects. Recorded on the audio channels were a number of control signals which allowed the computer to search any desired section of the tape and automatically play it back. Other audio signals made it possible for the computer to determine the time required to detect a target. The computer could also determine the correctness of the designation by comparing the output of the hand control with the known target position.

Subjects

Subjects were employees of the Aviation Research Laboratory at the University of Illinois. None of the subjects had previous experience with time-compressed displays, and all were unfamiliar with the hypothesis being tested.

Procedure

The procedure was very similar to that used by Scanlan, Roscoe, & Williges (1971). A notable exception was the use of a hand control and cursor to designate the detected target. This change was made after virtually identical results were obtained in a partial replication of the 1971 study using a hand control to designate the target.

Subjects were given 40 seconds to search after the first appearance of the target. If they failed to find the target in that time, a miss was recorded and a detection latency of 40 seconds was assigned. If something other than a target was identified, a false detection was scored and again a latency of 40 seconds was assigned.

Results

The average time to detect and the average number of missed or falsely detected targets for four trials and three subjects are given in Table 1. An examination of Table 1 indicates that in all three modes a playback rate of 100 milliseconds per frame yields better performance than a rate of 200 milliseconds per frame. These data also indicate that the time-compression/pause mode of presentation is not as effective as either of the other two modes. Finally, these data show no difference between the laydown and standard time-compression modes for the 100 millisecond per frame playback rate.

Discussion

The results of this limited study indicate that the procedure and variable levels used may not be adequate for testing the hypothesis

TABLE 1

Mean time, in Seconds, to Detect a Target
for Three Subjects, Three Modes of
Presentation, and Two Playback Rates

Playback Rate (milliseconds per frame)	Mode		
	Time- Compression	Time- Compression with Pause	Laydown
100	21.1 (0.67)	26.5 (1.67)	21.1 (0.67)
200	22.5 (0.67)	31.2 (2.67)	26.5 (1.00)

Note: Values shown in parenthesis are the average number of missed and falsely detected targets.

of improved performance with spatial and temporal cues. At least two aspects require further investigation. First is the question of task difficulty. For the time-compressed display the particular noise level selected produced only a moderately difficult task, and subjects required five or six seconds to detect a target after it was fully developed. If the laydown mode were considerably easier, this level of task difficulty may not be sensitive enough to indicate any difference. Additional study of these two modes needs to be conducted using either a higher noise level or a side task to increase the task difficulty.

Second is the question of search strategy. With the particular procedure used the playback rate or the mode was changed every four trials. This apparently created a problem for subjects who reported that they were just getting "tuned" to a particular playback rate and mode by the fourth trial. It is apparent from these comments that a larger number of practice trials are required prior to testing in a particular set of conditions. The problem of "tuning" did not appear in previous studies (Scanlan, et al., 1971) because 18 trials were given between rate changes.

References

- Scanlan, L. A. *Time-compressed displays for target detection*.
University of Illinois, Aviation Research Laboratory Technical
Report ARL-71-9/AFOSR-71-2, May 1971.
- Scanlan, L. A., Roscoe, S. N., & Williges, R. C. Time-compressed
displays for target detection. In S. N. Roscoe and B. H. Williges
(Eds.) *Three lessons in aviation research. Aviation Research
Monographs*, 1971, 1(3), 41-66.
- Wertheimer, M. Experimentelle Studien über das Schen von Bewegung.
Zeitschrift für Psychologie, 1912, 61, 161-265.

AN ASSESSMENT OF SYMBOLIC AREA

NAVIGATION DISPLAY VARIABLES

Richard S. Jensen

Aviation Research Laboratory, University of Illinois

Pilotage errors in area navigation were measured in flight for eight Airline Transport Pilots as functions of two navigation display variations--integrated heading versus separate heading presentation. The major task variable was the angle between successive route segments. The results indicated that horizontal and vertical steering errors are smaller than the values assumed by the FAA. There were no significant overall differences in pilotage errors for the two variations in the navigation display.

In 1969, with the release of Advisory Circular AC 90-45, the Federal Aviation Administration made a realistic approach to the problem of assigning protected airspace to aircraft flying in the United States national airspace system using area navigation equipment for navigation. Previously, pilotage errors were allowed by the assignment of very conservative buffer zones between one aircraft and another and ground obstacles. One of the most significant sections of the Advisory Circular is a total system error budget which includes both equipment and pilotage errors. In this new error budget certain magnitudes of pilotage error are assigned for given flight situations. Pilotage error is then combined mathematically with other sources of error in the system, resulting in a total system error for given flight situations. This result is then used to assign protected airspace to aircraft operating under given conditions. Before such a system can be implemented, it is necessary that data be collected and analyzed to establish empirical values for pilotage error as a function of major area navigation display design variables for representative classes of pilots under given flight situations.

The dependence of pilot performance upon flight display design has been demonstrated in many simulator and flight experiments (Bauerschmidt & Roscoe, 1960; Jensen & Roscoe, 1971). It has been hypothesized that one navigation display design variable which affects pilot performance is integrated presentation of heading and course deviation, as opposed to separate heading and course presentations (witness the large number of such displays in aircraft where cost is not a factor). It is argued that heading and course deviation information are directly related. Therefore, placing both in the same display should result in a corresponding decrease in tracking

error, because of reduced scan requirements. This difference can only be demonstrated in a flight task which is sufficiently difficult to require better display in task performance. The purpose of this experiment is (a) to determine, for a given group of pilots, the amount of pilotage error which should be assigned for terminal area navigation in level flight and (b) to determine whether an integrated heading and course presentation significantly improves pilot performance over a separate heading presentation.

Method

Subjects

The eight subjects for this study were Air Transport Pilots on current flight status for the Staff Air Transport Service at the University of Illinois, Urbana-Champaign. All had at least 3,000 hours of total flight time and at least 200 hours of instrument time. In addition, all subjects had participated in a previous area navigation experiment and were well acquainted with area navigation procedures.

Flight Facilities

The flight research facility used was a Beechcraft C-45H equipped with a Narco area navigation system and a Collins FD-109 flight director system. These two systems were integrated, permitting the selection by means of a switching panel of area navigation information on either the Narco Course Deviation Indicator (CDI) or the Collins Horizontal Situation Indicator (HSI). The scale factors on the two displays were made approximately equal.

Experimental Design

The experiment was designed so that the primary question, that of overall differences due to displays, would receive the most powerful test. A within-subject design was employed in which each pilot used each display the same number of times in a counterbalanced order on four flights over four courses. Each pilot flew two flights on one display, followed by two flights on the other display. These flights will hereafter be referred to as first and second trials. Four different courses were used, each composed of eight 15-nautical mile segments with four intercept angles: 22 degrees, 45 degrees, 67 degrees, and 90 degrees. Each angle occurred in both the first and second half of each course. Angles, courses, and flight were all counterbalanced with displays. The experiment was designed to test the effects due to displays in the following conditions: (a) overall conditions, (b) over the four turn angles, and (c) over the two trials on each display. The dependent variables were course crosstrack error, altitude error, and procedural error.

Procedure

Before flight the subject was told that the flight was to be made as a normal instrument flight in a terminal area. He was further told that he would receive clearances from the safety pilot as flight progressed, including sources to the next waypoint. He was given a chart of the local area which provided him with waypoint location and coordinate information but did not show the course he would fly. The experimental portion of the flight was made at 4,500 feet. The first clearance was given on ground. Thereafter, a new clearance was issued during new course interception following passage of each waypoint. Each successive waypoint required the setting of a new frequency which was to be set after passing mile seven outbound from the waypoint. There was a total of five procedural operations required on each 15-mile segment: (a) VOR frequency, (b) DME frequency, (c) waypoint radial, (d) waypoint distance, and (e) course selection. These procedural operations, in addition to the clearance which had to be copied on each segment, made the workload somewhat representative of workloads typical in terminal area flying. Crosstrack and altitude errors were recorded continuously during the flight on a two-channel strip-chart recorder. Procedural errors were recorded manually by the safety pilot and then brought to the attention of the subject.

Results

Pilot performance for each display was assessed in terms of the three types of error recorded: crosstrack, altitude, and procedural. Strip-chart recordings of crosstrack and altitude error were scored at one-mile intervals from five miles before the waypoint through seven miles after the waypoint. Procedural errors were recorded for all flights and summed together by type.

Crosstrack Error

Central tendency and 2σ variability for crosstrack error were determined for each display as a function of distance from the waypoint. The results showed no significant differences between displays for any of the four turn angles. In terms of absolute value the 2σ crosstrack variability, for both displays on second trials only, exceeded one mile only in the case of the 90-degree turn. After the new course had been captured, the crosstrack error appeared to be approaching ± 0.5 mile. These results indicate that the 2σ assumed crosstrack error of ± 1.0 mile by AC 90-45 is somewhat conservative.

An analysis of variance was done on the new course capture data from the interval 0 through 7 miles after waypoint. Because a significant amount of learning occurred from first to second trial ($p < .02$), only second trials are included in the data. Considering display differences over all conditions, the average RMS error for

the course capture interval was .591 nautical mile for the HSI and .640 for the CDI on second trials only. However, this difference was not significant. When angles are considered separately, there was a highly significant difference in performance between the small and large turn angles ($p < .0001$). The display differences, though not significant, do appear to be greater at the larger turn angles (see Figure 1).

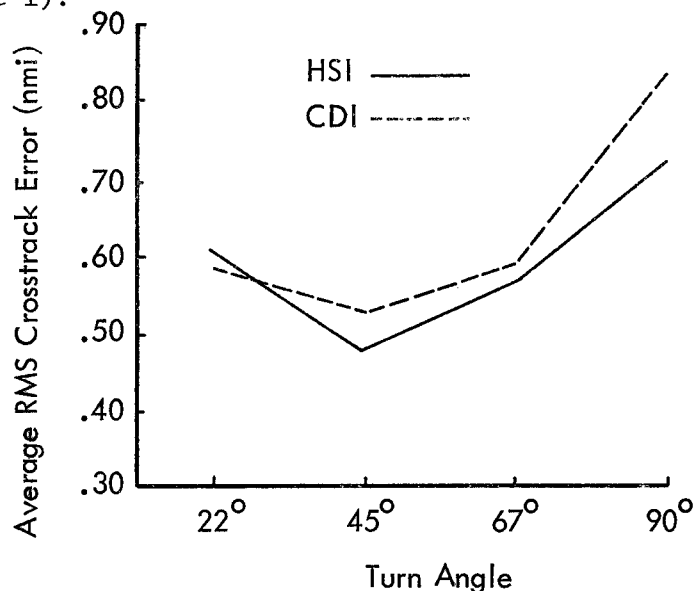


Fig. 1. 2RMS crosstrack error in nautical miles by turn angles and displays during course capture (0 - 7 n mi outbound) on second trials only.

Altitude Error

3RMS altitude errors were calculated for second trials on all segments and plotted as a function of distance from the waypoint (see Figure 2). These data indicate that the 3RMS vertical error was less than 100 feet for all second flight data. These data further indicate that vertical error was greatest near the waypoint while the turn and new course intercept were being made. In general these data suggest that altitude control performance was better when the CDI was being used than when the HSI was being used as the course reference. The greatest difference seemed to be during course capture. However, an analysis of variance showed no significant differences between displays.

Procedural Error

Procedural errors were scored and tabulated by type of error. Table 1 presents all errors of each type made during the experiment for the display being used at the time the error occurred.

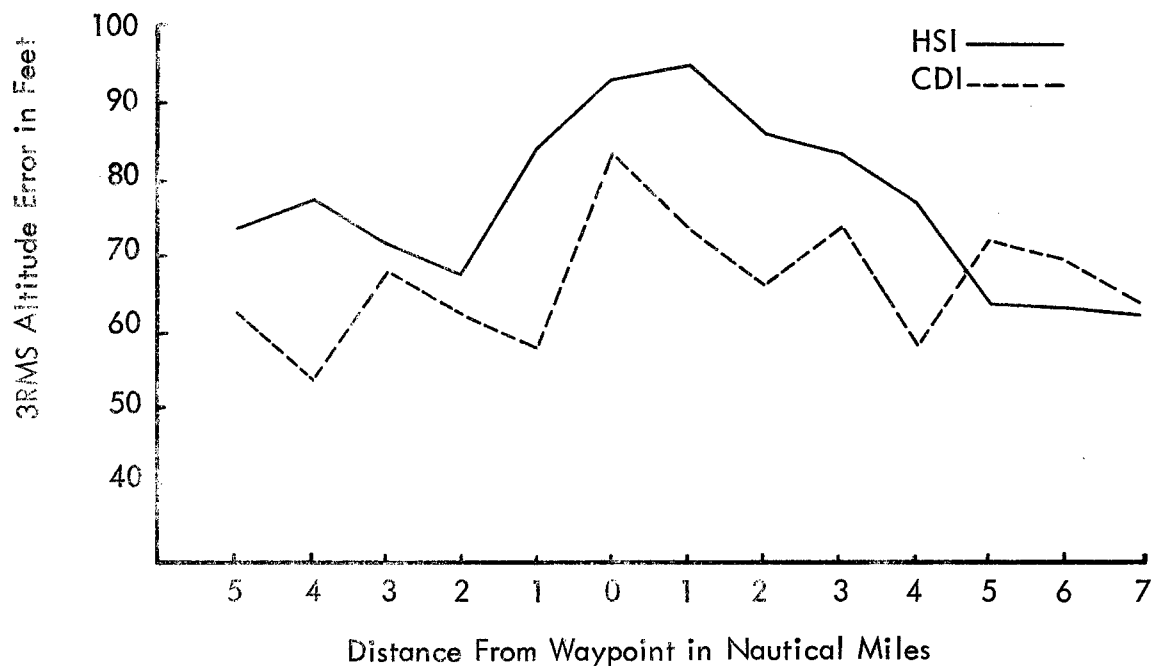


Fig. 2. 3RMS altitude error for all turns on second trials only.

TABLE 1
Procedural errors by Eight Pilots
Over Four Flights

Display	Frequency		Waypoint		Course Selector	Failure to Note Wpt	Total
	DME	VOR	Radial	Distance			
HSI	4	4	5	5	3	3	24
CDI	2	3	5	5	4	1	17

A total of 41 errors were made out of 1,536 procedural operations performed. Each number in Table 1 represents the number of errors out of 128 procedural operations performed of that type. 2.67 percent of all procedural operations resulted in a procedural error.

Discussion

Results from the three measured variables, crosstrack, altitude, and procedural error, clearly indicate that for a relatively easy two-dimensional task the experimental pilots were able to stay within the error limits set by FAA Advisory Circular AC 90-45 for terminal area operations even while making course changes of up to 67 degrees. Altitude error was particularly small in comparison with the tolerances established for vertically guided area navigation under similar conditions. One possible reason for these small errors is that all of the data was taken in atmospheric conditions where there was no perceptible turbulence.

Display Differences.

Data for the interval 0 to 7 miles after the waypoint show no significant differences in pilot performance for one display over the other. On second trials only there was slightly better crosstrack steering performance with the HSI than with the CDI for larger turn angles. However, on these same trials, altitude control was slightly better with the CDI display than with the HSI. These results seem to indicate that for the type of flight task used, the addition of integrated heading to the navigation display, although it is desirable from the pilot's point of view, does not improve his flying performance by a significant amount.

Procedural Error

The results for procedural error show that the possibility of procedural error exists any time a procedural operation is required. These errors occurred in similar frequencies for every required operation. A common error that contributed to frequency setting errors and waypoint coordinate setting errors was misreading the chart. A second common error was mistaking waypoint radial for course setting. A third common error was failure to notice the passage of a waypoint. This third error does occur with angular deviation but is more likely in linear deviation because there is no advanced warning on the display close to the station. It seems clear that the best way to reduce procedural errors is to reduce the number of procedures required in flight.

References

- Bauerschmidt, D. K., & Roscoe, S. N. A comparative evaluation of a pursuit moving-airplane steering display. *IRE Transactions in Human Factors in Electronics*, 1960, 1(2), 62-66.
- Federal Aviation Administration. *Approval of area navigation systems for use in the U. S. National Airspace System*. Washington, D. C. Government Printing Office, FAA Advisory Circular AC 90-45, August 1969.
- Jensen, R. S., & Roscoe, S. N. *Performance of instrument pilots using symbolic area navigation displays*. University of Illinois Aviation Research Laboratory, Technical Report ARL-71-23/FAA-71-3, October 1971.

EVIDENCE FOR A PROCESS MODEL OF INTELLIGENCE ¹

Earl B. Hunt and Nancy Frost ²

University of Washington

Simulations of man/machine systems require explicit models of human performance. Such models are used in modern theories of cognition. The models make no provision for individual differences. Wide individual differences, however, are found in performance in real-time computer applications. We have established a correlation between the parameter values for information processing and scores on conventional aptitude tests. Short term memory is associated with verbal aptitude and intermediate term memory with quantitative aptitude. These results can be used to simulate the performance of a given individual in a specific man/computer task. The work is being extended to decision-making and perceptual abilities.

A few years ago the White House asked if it were possible to predict the responses of a specific individual in a number of hypothetical situations. The individual was Nikita Khrushchev, and the need to predict his behavior has passed (Frederiksen, 1972). The scientific challenge remains. We often need to predict how individuals will react in situations which cannot be tested before it is too late. While personality factors are important, we will be concerned with predictions based on information processing capacity. We feel that modern intelligence tests, including personnel classification batteries, are incapable of meeting this challenge. Our reasons are theoretical rather than empirical; the tests were never designed to make the type of prediction that is needed. We shall propose an alternative and present preliminary evidence suggesting that our alternative is at least worth exploration, and then indicate how the alternative could be used.

Binet introduced the idea of defining intelligence by comparing individuals. Since then our measures have become more sophisticated, but multidimensional comparisons are still comparisons. An intelligence test only inferentially tells us something about the absolute level of performance of an individual, or about the processes which he uses to solve problems. Consider an analogy to automobiles. We could develop a factorial test of automotive performance based on inter-car comparisons, but we do not, because we know how a car works. We can better predict the performance of a particular model by stating a few parameters describing its components. Our point is that the same thing is possible in describing mental capacity. This is hardly

a new idea. Francis Galton tried 100 years ago with a bad model. Today things *may* be better. Within the past ten years psychologists have developed and experimentally verified an apparently workable general model of cognition based largely on analogies to information processing in computer systems. As is the tradition in experimental psychology, the appearance of wide individual differences in the data has been regarded as a noisy nuisance, and either controlled or neutralized by statistics. It is not surprising that modern cognition has had little influence on personnel selection! Our long range goal is to change this situation. We believe that we can return to Galton's original idea of a process model of intelligence, and use it to develop a test of a person's absolute performance level. In this presentation we shall provide evidence that the first step can be taken along the path without falling down.

36

A General Information Processing Model

The experimental model with which we have worked is a variation of a buffered memory model which appears to encompass most of the recent experiments (Hunt, 1971). The model in outline is shown in Figure 1. It assumes that there are three separate stages of memory: a *sensory* or *iconic buffer memory* where very fleeting and perception-like representations occur, a *conscious memory* which includes data transferred from the iconic buffer by selective attention processes and in which data comparisons can take place, and a more permanent *long term memory*. The last probably consists of two components, an intermediate memory of our experience over the past few minutes and a dictionary of all our past experience. The basic parameters inferred from this model describe the size of each buffer, the rates of information transfer between them, and the speed and accuracy of elementary data processing operations within each of the stages of memory. Comparisons and search operations are of special importance. There are a number of micro-models which can be used within this framework to analyze experiments testing each parameter. These provide the basis for our data.

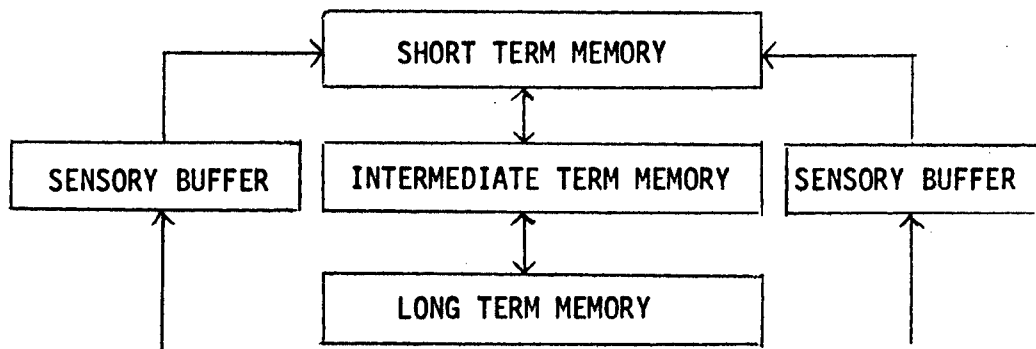


Fig. 1. A computer-like model of memory, from Hunt (1971).

The Experimental Plan

To establish the existence of reliable differences in information processing, we decided to try to relate individual parameter values to the two most reliable measures of individual differences in intellectual performance, quantitative ability (QA) and verbal ability (VA). If the hypothetical "new test" approach is feasible there ought to be some correlation to old tests, since the old tests are useful in many situations. Ideally, we would have selected information processing in subjects from extreme ends of the QA and VA dimensions. Practical considerations made this difficult. We had to be content with studying extremes of ability within the available subject population: college undergraduates at the University of Washington. A panel of subjects was recruited from students in the four combinations of top and bottom quartile VA or QA within the freshman class. The Washington Precollege test battery was used to establish VA and QA scores for our subjects. Thus our "high" and "low" subjects were defined relative to a college population and not the general population of young adults. Note, however, that our population probably is intellectually comparable to the officer corps.

Experimental Studies

The first experiment used a paradigm devised by Atkinson and Shiffrin (1968), in which subjects perform a difficult paired associates learning task. The subject observes a continually varying sequence of CVD-digit pairs, with interspersed test and study trials. The information processing requirements of this task are similar to those placed on a radio operator who must keep track of the current location of several aircraft when the aircraft are continually sending new position reports. A mathematical model for analyzing this task yields the following parameters:

**conscious memory capacity (r)*

**probability of transfer from the sensory buffer to conscious memory, an attention parameter (α)*

**rate of transfer from immediate conscious memory to the intermediate memory store (θ)*

**decay rate from the intermediate memory store (τ).*

We found that subjects with high QA scores showed significantly better retention of information over a period of five minutes, i.e., they have a lower value of τ . Only insignificant differences between individuals were found in the rate of transfer of information from conscious memory to the intermediate memory store. There were large individual differences in the size of the immediate memory buffer

and in the attention parameter, but these did not appear to be related to VA or QA scores, at least within the restricted range of our sample. It is of interest to note that Loftus (1971) has evidence indicating that these parameters may be varied by the subject's strategy.

A second study concentrated on tasks we intuitively associate with VA. The data indicated that the association exists, but not in the way we expected. Twelve in each of the four combinations of high and low QA and VA performed a free recall task using categorized lists. In one condition the subjects viewed 30 words presented in blocked sequences: e.g., animals, then vegetables, then minerals. In another condition the stimulus items could, indeed, be categorized into groups, but were presented in random order. By noting the order of free recall, we could determine whether the subjects were likely to recall items by category (semantic clustering) or by rote, in the order in which the items were presented.

In the case of the blocked lists there was no difference, and all subjects recalled a high number of items. There was a definite change in strategy in the random list condition. The high VA subjects displayed *less* semantic clustering than the low VA subjects. The data suggested to us that the high VA subjects were better able to hold verbal items in their STM buffers than were low VA subjects. This idea was corroborated by the subjects' reports: high VA subjects said they simply read back the words in the list, while low VA subjects reported systematic search strategies.

While this explanation is a reasonable one, the evidence and the explanation have an *ad hoc* air. A third experiment sharpened the picture considerably. If it is true that the high VA subjects are better able to manipulate information in STM, then they ought to be able to make comparisons between items of information in STM more rapidly than low VA subjects. After all, this ability is one that would be particularly important in speech comprehension, something that high VA subjects are good at by definition. This prediction can be tested directly by using Sternberg's (1966) paradigm in which a subject must remember a set of 1 to 5 digits. The digit set (called a *memory set*) is displayed visually. Seconds later a probe signal is presented and the subject reports whether the probe was a member of the memory set. It has been found that the reaction time (RT) in this situation is a linear function of λ , the size of the memory set. The slope of the RT as a function of λ is interpreted as an estimate of the time required for the comparison of two characters in memory. We have found that on the average, high VA subjects can make this comparison almost twice as fast as low VA subjects. The appropriate comparison is shown in Figure 2.

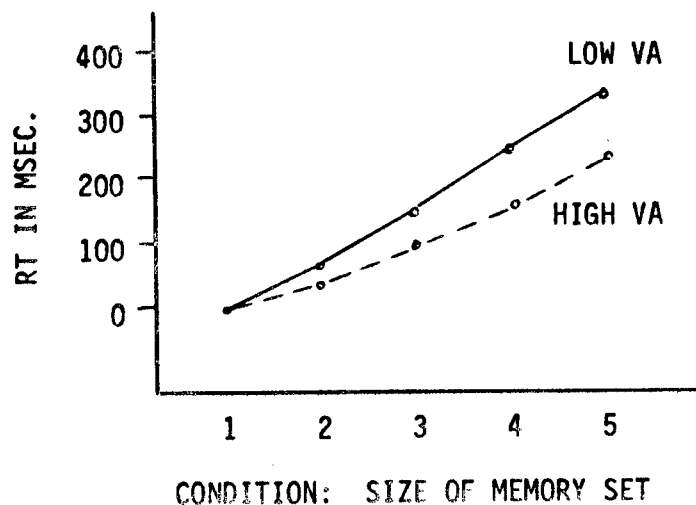


Fig. 2. RT to correctly recognized probes from the memory set. Data points represent the mean of RTs at each condition minus the mean RT for Condition 1. Curves for individual subjects were smoothed.

Finally, we must report a negative finding. Many tasks require that people hold data in STM while performing a possibly interfering distractor task, and an alternative explanation of the semantic clustering results might be that high VA and low VA subjects differ in susceptibility to STM interference. We have repeated the Peterson & Peterson (1959) "counting backward" paradigm and found very great individual differences in susceptibility to retroactive interference. At the extreme we have found one individual who is apparently impervious to interference in STM. (This individual has a remarkable memory in other ways as well. The details of his performance have been reported elsewhere (Hunt & Love, 1972). On the other hand, we have found no correlation whatsoever between resistance to interference in this task and either VA or QA. Apparently the tests do not measure this ability.

Conclusions and Prospectus

Table 1 summarizes the experimental findings. We have shown that high QA is associated with an ability to handle items in intermediate term memory and high VA with an ability to manipulate STM. There is evidence that the ability to manipulate STM may lead individuals to over-rely on STM, as apparently happened in the clustering study. We are now extending our work, both by examining a larger panel with subjects varying along perceptual as well as QA and VA dimensions, and by conducting experiments in decision-making and attention to supplement our studies of memory. Rather than speculate about how these studies will come out, we would like to go a step further and ask what would happen if they do.

Why does society, and military society in particular, need a new intelligence test? In our introduction we indicated the most important reason, the need to predict absolute performance in specific situations. Typically we will be concerned with the performance of a man/machine team rather than the performance of man alone. It is usually possible to increase system performance either by selecting better people or by building better machines. Which route to take depends on the costs and benefits expected from each combination of ability and machine quality.

TABLE 1

Summary of Experimental Findings

Phenomenon	Large Individual Variation	Association with IQ
STM Size	Yes	No
Attention	Yes	No
Transfer to ITM	No	No
ITM Decay	Yes	QA
Large STM for Verbal Data	Yes	VA
Speed of Comparison of Names in STM	Yes	VA
Susceptible to Interference	Yes	No

To establish expectations we often resort to simulation, especially when actual exercise of the system is expensive or impossible. To represent the machine part of the system being simulated we write a computer program whose logic is dictated by the interaction between machine components and whose parameters are determined by our knowledge of component performance. To date, the human part of the system has been represented by a set of task-specific parameters obtained either by special experiments or by guesses based on tangentially related publications about humanity in general. We would improve upon this by providing a basic model of man as a computing system. To represent human performance, simulation designers should consider the program needed for the human computing system to accomplish its task. To represent an individual, the designer should

insert into the basic model of man parameters representing the information processing capacity of the person in question. The necessary task-specific parameters would then be calculated by inference.

To reach our Nirvana of simulation we need three things: a viable general model of man; evidence that there are reliable differences in information processing parameters; and a testing technology for establishing these parameters at reasonable cost. The theoretical work on the model is well under way; we have here reported the very first steps in the data gathering. We are confident that the appropriate technology can be developed, although it may well rely on interactive computing rather than on traditional paper and pencil methods of test administration.

Footnotes

¹The research reported here was sponsored by the Air Force Office of Scientific Research, Air Force Systems Command under grant 70-1944 to the University of Washington.

²We would like to thank Professor Clifford Lunneborg for his assistance throughout, and in particular for aid in selecting the panel of subjects described here. Our thanks also go to Benôit Côté, Michael Irrgang, Phillip Milliman and Susan Nix for their assistance in this research.

References

- Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The Psychology of Learning and Motivation*, Vol. 2. New York: Academic Press, 1968.
- Frederiksen, N. Toward a taxonomy of situations. *American Psychologist*, 1972, 27, 114-123.
- Hunt, E. B. What kind of computer is man? *Cognitive Psychology*, 1971, 2, 58-98.
- Hunt, E. B., & Love, L. T. *How good can memory be?* University of Washington Technical Report. Proceedings of the 1971 NAS-NRC Conference of Coding in Memory, 1972.
- Loftus, G. R. Comparison of recognition and recall in a continuous memory task. *Journal of Experimental Psychology*, 1971, 91, 220-226

Peterson, L. R., & Peterson, M. J. Short term retention of individual items. *Journal of Experimental Psychology*, 1959, 58, 193-198.

Sternberg, S. High speed scanning in human memory. *Science*, 1966, 153, 652-654.

Human Factors

CHAIRMAN: Captain Richard L. Hughes

PERCEIVED DESIRABILITY OF ASSIGNMENT AS A
PERSONNEL SUBSYSTEM MANAGER

William H. Hendrix

Air Force Systems Command

The study was conducted in order to establish the desirability or undesirability of assignment as a Personnel Subsystem Manager. Research data were collected on Personnel Subsystem Managers and non-Personnel Subsystem Managers assigned to the Air Force Systems Commands; Electronic Systems Division, by use of a questionnaire. The perceived desirability of a Personnel Subsystem Manager Assignment, the status of the Personnel Subsystem functional area, and the value of Personnel Subsystem Manager personnel are tabulated and discussed.

The total number of Personnel Subsystem Managers (PSMs) within the Air Force Systems Commands; Electronic Systems Division (ESD) has been on the decline. In addition, there is a lack of well trained PSMs available for assignment to Program Offices (POs).

The purpose of this study was to determine whether the perception of the Personnel Subsystem (PS) career area held by PO personnel has contributed to these conditions. More specifically: (a) does the PSM perceive his job as desirable or undesirable, and (b) how does the non-Personnel Subsystem Manager, located within ESD, perceive the Personnel Subsystem (PS) area.

The non-Personnel Subsystem Manager (non-PSM) group required attention because: (a) the group provided a source from which future PSMs could be developed, and (b) their perception of the PS area could influence both the status of the PS area within the PO, as well as the view held by PSMs assigned to POs.

Air Force Regulation 80-46 establishes the requirements for assignment as a PS Manager. The basic requirement being that an individual must possess an Air Force Specialty Code of either 2955 (Personnel Subsystem Manager) or 2675A (Human Performance Engineer).

Method

The research data were collected by use of a Personnel Subsystem Questionnaire. The questionnaire was designed to obtain information associated with the perception of the PS area by ESD personnel. Items on the questionnaire consisted of two types: (a) biographical items,

and (b) rank-order ratings.

The rank-order ratings focused on two major areas. One area was the main career or job areas to which individuals within the PO are assigned. These career areas, identified by Air Force Specialty Codes (AFSCs) were: (a) 2675A - Human Performance Engineer, (b) 2825 - Electronics Engineer, (c) 2835 - Mechanical Engineer, (d) 2935 - System Program Data Management Officer, (e) 2955 - Personnel Subsystem Manager, (f) 3055 - Communication-Electronics Engineer, and (g) 5125 - Computer Systems Design Engineer.

The other major area dealt with the prime functional areas which are subcategories under the overall Management process. These functional areas are listed in Table 3.

The questionnaires were administered to 26 Ss assigned to ESD. This group was composed of two subgroups of 13 Ss each. One group consisted of all known PSMs assigned to ESD (PSM Group). The other group consisted of engineering personnel who had never served as PSMs (Non-PSM Group). Individuals assigned to the Non-PSM Group were obtained by a stratified sample from the five major procuring organizations (Deputies) at ESD.

The data collected were tabulated. For each rank-order item, a mean across all Ss for each group (PSM Group and Non-PSM Group) was computed. Based on the mean values a PSM and Non-PSM Group rank-order listing was derived (Tables 1, 2, and 3).

Results

PSM respondents indicated that a 2955 - Personnel Subsystem Manager, was more beneficial to a Program Office than was a 2675A - Human Performance Engineer, for managing the PS program. 84.6% responded in favor of a 2955 - PSM. The Non-PSM Group also indicated this preference with 76.9% preferring a 2955 - PSM.

When asked if a PSM (2675A or 2955) was required to perform the PS Management tasks, or could the PS tasks be performed equally well by engineering personnel, both groups indicated that an assigned PSM was preferable. The responses in favor of a PSM were 84.6% by the PSM group and 61.5% by the Non-PSM group.

Rank-order ratings of AFSCs in order of importance to the Program Office (PO) revealed differences between the PSM and Non-PSM groups (Table 1). The PSM group indicated that a 2955 - PSM was one of the most important types of individuals who could be assigned to PO (2.5/7). The Non-PSM group on the other hand indicated that a 2955 - PSM was one of the least important individuals (5/7).

TABLE 1

Relative Importance of AFSCs to Program Offices

AFSC	PS Managers		Non-PS Managers	
	Rank Order	Mean	Rank Order	Mean
2825	1	2.25	1	1.31
2955	2.5	3.67	5	4.92
5125	2.5	3.67	3	4.08
3055	4	4.25	2	3.00
2675A	5	4.42	6	5.15
2835	6	4.58	4	4.00
2935	7	5.17	7	5.54

Rank-order ratings listed with the rating first followed by the total rating categories (2.57) means a rank-order rating of 2.5 (tied for second and third place) out of 7 possible categories.

The rank-order ratings on best career AFSC (Table 2) indicated agreement between the PSM and Non-PSM group when the ratings were divided into upper (rank-order 1-3) and lower (rank-order 4-7) groups. Both groups rated the 2955 - PSM and 2675A - Human Performance Engineer AFSCs in the lower group.

Rank-order ratings for functional areas (Table 3) revealed that PSMs rated the PS functional area approximately mid-scale in importance to the PO (5/8). The Non-PSM, on the other hand, indicated it was the least important of all functional areas (8/8).

Discussion

A previous study (Hendrix, 1971) conducted at ESD found that PS deficiencies were in the main due to the low status of the Personnel Subsystem area within the program office. The present study tends to support that conclusion. Non-PSMs rated the PS functional area as the least important of all areas (8/8), and one of the worst areas for a career [2955 (5/7), 2675A (6.5/7)]. They also indicated that a PSM is one of the least important individuals to be assigned to the Program Office [2955 (5/7), 2675A (6/7)].

TABLE 2
Best Career AFSCs

AFSC	PS Managers		Non-PS Managers	
	Rank Order	Mean	Rank Order	Mean
2825	1	1.39	2	1.33
5125	2	1.46	1	1.25
3055	3	1.77	3	2.00
2675A	4.5	2.15	6.5	2.50
2835	4.5	2.15	4	2.08
2955	6	2.31	5	2.25
2935	7	2.77	6.5	2.50

TABLE 3
Functional Areas

Functional Area	PS Managers		Non-PS Managers	
	Rank Order	Mean	Rank Order	Mean
System Engineering	1	1.00	1	1.31
Reliability & Maintainability	2	3.33	3	4.00
Test and Evaluation	3	3.92	2	2.46
Computer Programming Management	4	4.50	4	4.62
Personnel Subsystems	5	4.75	8	6.23
Quality Assurance	6	5.17	7	6.15
Configuration Mgmt	7	6.08	5	5.54
Data Management	8	7.25	6	5.69

8 The PSM group, on the other hand, perceives the 2955 - PSM as one of the most important individuals supporting the PO (2.5/7), and yet one of the worst career areas in which to be assigned (6/7).

Conclusion

It is concluded that the Personnel Subsystem area at ESD continues to occupy a position of low status within the Program Office, and the PSM career area is perceived as less desirable than most other Program Office associated career areas.

Reference

Hendrix, W. H. Isolation of personnel subsystem management problem areas. *Proceedings of Second Annual Symposium Psychology in the Air Force*, USAF Academy, April 1971.

ARMY AIRCRAFT SURVIVABILITY HUMAN FACTORS

Robert W. Bauer

Human Engineering Laboratories, United States Army

A comprehensive program of research on the human factors in combat aircraft survivability is discussed, using the sequences of events in the actual air-ground engagements.

In 1970 and 1971 the Army Materiel Systems Analysis agency (AMSAA) and the Human Engineering Laboratory (HEL) at Aberdeen Proving Ground were engaged in a comprehensive program of research on the human factors in aircraft survivability. Before sending data collection teams to Vietnam in December of 1970, we worked through a considerable body of combat damage data which flowed into AMSAA in 1969 and 1970. As you may know, there are ten or more different reporting systems used for the collection of battle damage data from Vietnam and there is considerable duplication among these. Despite the masses of data, it was disturbing to discover that no single reference contains all of the Army aircraft hits! Furthermore, the Army, Navy and Air Force data are not recorded in a comparable form, so that the interpretation of comparisons at the DOD level must become extremely difficult if not impossible! If there is an advantage in this confusion of reporting methods, it cannot be an advantage for the executive levels in national defense.

I think it is very important for the largest Air Force in the world (USAF) and the third largest (USA) to get together on a uniform reporting system for combat damage which will permit rational comparisons of combat damage rates. We found the OPREP 5 data, the BDARP (Battle Damage Assessment Reporting Program) data and the Aircraft Inventory Status and Flying Time (AIS & FT) data most useful for our purposes. In the data we examined, reconnaissance missions and assault missions always accounted for the largest incidence of hits with fire missions generally in third rank in the tolls. With regard to aircraft taking hits, the largest incidence was among troop transports (UH-1D/Hs), second largest, scouts (predominantly OH-6As) with gunships (UH-1B/C/Ms and AH-1Gs) taking third place. The incidence figures on hit aircraft don't tell us which aircraft is more likely to be hit because they don't tell us which aircraft is flown more. In order to know more about attrition, we need an index which relates a hit incident to flight hours by aircraft, by mission, and by severity of damage.

Mission information appears in the OPREP 5 but is not linked to flight hours in a way that will permit calculation of combat damage risk by mission. However, we had (90% or better of) hit incidents (aircraft receiving one or more hits) from the OPREP 5 and we got

TABLE 1

Combat Damage in All Vietnam - 1969

3

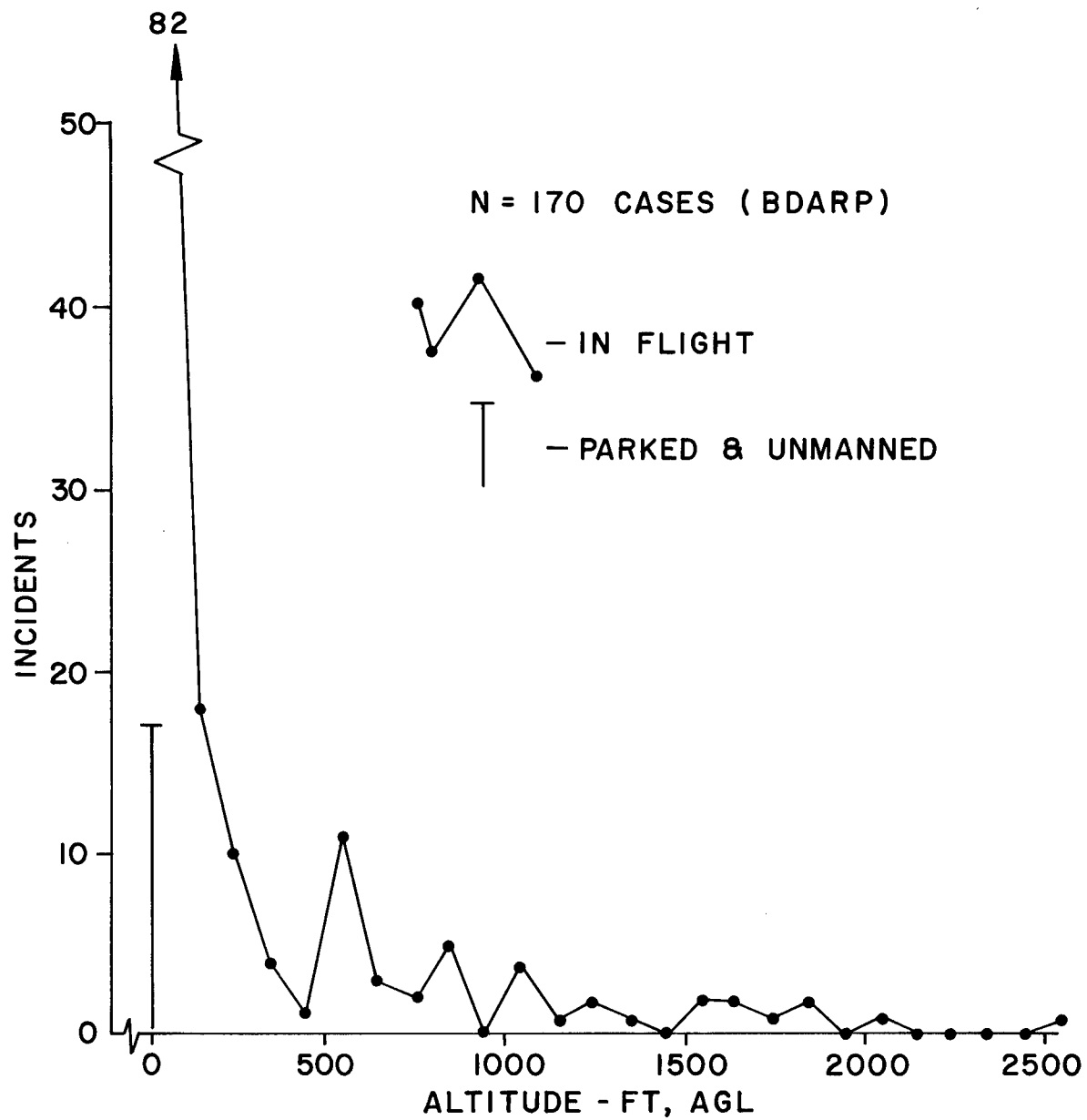
Type Aircraft	Mean Monthly No. of Damage Incidents (A+B+C+D)	Mean Monthly Flight Time	Mean Combat Damage Rate
OH-6A	119.1	36,366.3	.00328
UH-1B/C	105.1	19,096.0	.00550
UH-1D/H	223.9	135,023.5	.00173
AH-1G	63.5	25,992.3	.00244
CH-47	30.7	18,591.5	.00165
UH-1	339.0	154,119.5	.00219
All			

flight hours from the aircraft inventory status and flying time (AIS & FT). This permitted us to calculate combat damage rates patterned after the accident rate. It can be expressed as rate per hour as shown here or as rate per 100,000 hours, which is customary with accident rates. For example, the combat damage rate for all OH-6As in all of Vietnam in 1969 was 328 per 100,000 flight hours.

The Army has an objective method of classifying severity of damage, as indicated by the A, B, C, D on top of the second column. I won't go into it in detail here, except to say that this permits us to separate those aircraft which are lost or destroyed (D-damage) or those aircraft which can be repaired in less than 24 hours (A-damage). Seventy percent of all Army aircraft combat damage in 1969 was classified A or B (i.e., the estimated repair time was seven days or less.

Altitude and phase of flight showed some interesting results. Figure 1 represents 170 cases of aircraft hit reported through BDARP in 1969 and 1970. Ninety percent occurred at less than 1000 feet AGL and over 50% at less than 100 feet. The high proportion of hits at extremely low altitudes is related to the ground support role of Army aircraft and to the NVA's reliance on small arms, mostly AK47. I understand that about 70% of Seventh Air Force aircraft hits were also taken at less than 1000 feet AGL in South Vietnam, 1968 and 1969. But, altitude doesn't explain anything unless we can relate it to the

HIT INCIDENTS VERSUS ALTITUDE



aircraft and the mission. The OH-6A had a relatively high combat damage rate (328/100,000 flight hours) and almost half of these OH-6As were hit at a very low altitude and slow speed within the "dead man's curve," a flight regime in which autorotation is impossible. The AH-1G had a lower combat damage rate (244/100,000 flight hours). A very small percent (1%) were hit within the "dead man's curve." Ninety-nine percent were hit while flying over 50 knots and about 80% over 200 feet AGL). The gunships and OH-6As had some special problems in the application of suppressive fire. The suppressive fire was considered effective in reducing or stopping enemy ground fire in about half the cases in which it was applied, but one or more weapons jammed or were found inoperative during or just prior to the engagement in one case out of five! The AH-1G was relatively effective with suppressive fire but suffered an excessively high proportion of weapons malfunctions.

Enemy fire was detected by hearing in 70-80% of instances. Enemy fire was less often suppressed in hit cases, as one would expect. But among hit cases, over 70% received fire before the enemy was detected. Calculations indicated that probability of hit was 4X greater for aircrews who received fire prior to detection of the enemy before fire was exchanged. Aircrews of hit aircraft less often located the enemy. The predominant engagement sequence was: (a) received fire, (b) located enemy, and (c) returned fire. Aircrews of hit aircraft detected the enemy before fire was exchanged in only 17% of the cases, while aircrews of aircraft which escaped hits detected the enemy first in 33% of the cases.

As a result of our studies of about 180 BDARP cases from 1969-1970, about 7000 OPREP 5 cases from 1969 and about 400 cases directly interviewed in Vietnam in 1970 and 1971, we developed a number of recommendations on training, maintenance, design/development and reporting of hits.

With regard to training, we are now convinced that a much higher proportion of Army aviators must be IFR qualified. This will be consistent with the more advanced aircraft in development for the Army's and the increased IFR capability will have a favorable impact on both combat damage rates and accident rates. With the advantage of hindsight, we can see that the Army has neglected the training of combat crews *as crews*. The rapid turnover and replacement of crew members in Vietnam hampered the development of coordination and teamwork among aircrew members. In our airmobile operations, it is also becoming increasingly apparent that officers in the combat arms (armor, infantry, artillery) must become more familiar with aviation procedures and tactics. We foresee a need for expanded integrated training of airlift crews with ground troops and equipment.

We also foresee a need for improved unit maintenance facilities for more advanced aircraft, such as the Cheyenne. Maintenance personnel must be trained in CONUS to support the more complex aircraft systems and armament systems on the way--in development.

With regard to new designs, it is clear that increased power, lift capability and armor protection for critical components (including crew members) are needed. Furthermore, we are interested in determining if the future scout (light observation helicopter) should have a tandem seating arrangement rather than the side-by-side seating in the current LOH.

With regard to combat damage reporting I have already mentioned the need for more complete reporting. The dozen or more army systems could be efficiently consolidated. But, more important, is the need for a comprehensive joint services air combat reporting system. Such a system must include an acceptable scheme for classifying missions and roles of aircraft. It must include:

1. Flight hours (by aircraft, by mission, by role).
2. Aircraft type.
3. Role of aircraft.
4. Mission of aircraft.
5. Severity of combat damage (A, B, C, D).

With this information, a combat damage ratio

$$\text{COMBAT DAMAGE RATIO} = \frac{\text{NO. AIRCRAFT HIT}}{\text{NO. FH FLOWN}}$$

can be related to mission, role, and aircraft type.

References

- Bauer, R. W., Cassatt, R. K., & Waugh, J. D. *Air-Ground engagements in Vietnam* (U). (Study I - Interviews on aircraft survivability) (U). Technical Memorandum 10-71, Human Engineering Laboratories, Aberdeen Proving Ground, Maryland, August 1971. CONFIDENTIAL
- Cassatt, R. K., & Barnes, J. A. *Operational analysis of airmobile operations in South Vietnam during 1969* (U). (Study II - Analysis of high-damage units) (U). Technical Memorandum 5-71, Human Engineering Laboratories, Aberdeen Proving Ground, Maryland, April 1971. CONFIDENTIAL
- Cassatt, R. K., & Bauer, R. W. *Aircraft suppressive fire effects and weapons operational problems* (U). (Preliminary results of Vietnam data collection), Letter Report No. 138, Human Engineering Laboratories, Aberdeen Proving Ground, Maryland, July 1971. CONFIDENTIAL
- Department of the Army, Memorandum for the Director, DIA, Subject: Request for Intelligence Production Support (U). Office of the Assistant Chief of Staff of Intelligence, 27 February 1970. CONFIDENTIAL
- Waugh, J. D., Bauer, R. W., & Tucker, K. *Human factors in aircraft survivability*. Analysis of battle damage assessment reporting program data, Technical Memorandum 28-70, Human Engineering Laboratories, Aberdeen Proving Ground, Maryland, November 1970.

A PSYCHOLOGIST'S INPUT TO OPERATIONS RESEARCH

Gerald P. Chubb

Air Force Systems Command

The Systems Effectiveness Branch of the Aerospace Medical Research Laboratory is developing several classes of Monte Carlo models to reflect the impact human performance can have on mission success under a variety of mission conditions. Initial efforts modified the Siegel-Wolf two-man model and applied it to the F-106, assessing the impact of man-machine vulnerabilities to nuclear weapons effects. This model is being upgraded and applied to the B-52 prior to performing a proposed design evaluation of the B-1. A multi-man team performance model is also being developed to treat the information processing and decision making tasks associated with air surveillance and command/control systems.

The Air Force Weapons Laboratory (AFWL) has the responsibility for assessing the vulnerability/survivability of existing aerospace systems under nuclear attack conditions and for specifying hardness criteria for the design of future systems. In support of AFWL efforts, the Aerospace Medical Division (AMD) has organized an integrated program to treat human vulnerabilities to nuclear weapons effects. The USAF School of Aerospace Medicine (SAM) provides the radiobiology expertise, estimating human performance degradation postirradiation based upon clinical data from radiotherapy, nuclear accident data, and extensive research results using primates to study the effects of supralethal doses. The Aerospace Medical Research Laboratory (AMRL) provides the modelling expertise, integrating task data obtained from the operating command with the environment and hardware degradation data from AFWL and the human degradation data from SAM.

Rationale

While current activities focus on vulnerability/survivability assessment, AMRL's modelling efforts are actually oriented toward a more basic issue. The psychologist's role in human engineering has traditionally been more as a critique of design than as design engineer per se. Although great emphasis has been placed on incorporating human factors considerations into systems design at the earliest stage possible, to avoid the high expense of retrofit or redesign, human engineering techniques have not emphasized methods applicable in concept formulation, in the operations and systems analyses that define the operating requirements for future aerospace systems. More typically, the tools of the trade are geared to

subsequent stages of design, taking the systems operational requirements and ancillary mission constraints as virtually unalterable restrictions.

Background

In reviewing over 15 years of military operations research within the Department of Defense, ARINC Research Corporation, under contract to AMRL, found that human factors considerations were either tacitly ignored in these study efforts or were superficially treated in terms of gross assumptions about human performance. No major attempt was made to reflect the impact human performance, individual differences, training, or other considerations might affect systems performance and in turn mission success. This observation led to the definition of a need to develop man-machine models which could explicitly treat the impact human factors can (and do) have on mission success.

Approach

Studies to Date

AMRL's preliminary efforts in this area focused on acquiring a first-hand working knowledge of the Siegel-Wolf two-man operator simulation model (Siegel & Wolf, 1969). The initial effort was multipurposed (Siegel, Wolf, Fischl, Miehle & Chubb, 1971) and provided the background experience necessary to define a workable approach for treating human vulnerability to nuclear weapons effects.

At the request of AFWL, the Air Defense Command provided a block diagram representation of the F-106 intercept mission (Chubb, 1971) and a description of the tasks executed by the pilot. Based on the contents of the flight manual and tactics manual, AFWL further refined the analysis of pilot tasks. Applied Psychological Services, Inc., under contract to AMRL, then performed more detailed analyses in preparing the necessary input for the simulation model. The data were based upon interviews of experienced pilots of the 95th Fighter Interceptor Squadron at Dover Air Force Base, Delaware, including observations made using the ground-based flight simulator. Model runs using these data established baseline predictions of mission success for the selected intercept profile.

Applied Psychological Services also performed a literature search, identifying the performance implications of nuclear weapons effects. From this literature, quantitative descriptions were derived estimating the expected performance degradation postirradiation. Subsequent review of the results of initial simulation runs led to changes and refinements in the description of radiation induced performance decrements. The methodological approach and the decrement curves are also documented in Chubb (1971). More extensive documentation of the

evolution of the modelling effort is still in preparation but will ultimately appear as an AMRL Technical Report.

In the first effort, the analysis tacitly assumed: (a) a short, "worst case" intercept situation, (b) mission completion at the launch of all stores, and (c) fully functional hardware. It was recommended that in follow-on efforts these assumptions be relaxed. Laboratory director funds were then obtained to: (a) extend the task data to include other functions (e.g., CAP - Combat Air Patrol, RTB - Return to Base, etc.); (b) segment the mission by phase (and allow separate consideration of each attempted pass); and (c) develop an extended capability to treat the impact hardware vulnerabilities would have on pilot performance, effectively integrating a joint consideration of both human and hardware degradation.

The extension of the task data and the model modifications required to segment time stress by mission phase were both straightforward and will not be treated here. Documentation is being prepared in an AMRL Technical Report. Of perhaps greater interest is the approach taken to integrate human and hardware vulnerability estimates.

Hardware degradation due to any cause (designed reliability of the equipment, conventional weapon battle damage, or nuclear-induced malfunction) will reflect itself as a change in the operating performance of controls and/or displays. The change may be permanent or transient with obvious or subtle symptoms of operating deficiencies, and the deficiency may be either "all or none" in nature or may be quantifiable as a magnitude change in uncertainty. Each of these considerations has implications for the modeller/analyst. As a first attempt, it was assumed that radiation effects on the equipment became manifest as an obvious, permanent and complete cessation in some display or communication device, which AFWL believed was appropriate for the F-106 but may not be for other systems.

In this restricted case, AFWL provided curves which showed the probability that a functional subsystem would cease to operate as a function of neutrons per square centimeter. Equipment (radios, displays, etc.) associated with these functions would be therefore "unavailable" to the pilot after radiation-induced malfunction, and the pilot would be obliged to alter the tasks performed to complete the intercept. Given the combinations of equipment malfunctions that could occur, experienced pilots were again interviewed to determine what actions they would take if these conditions arose. The problems with this approach and some proposed alternatives are explored in Chubb (1971).

The simulation of system vulnerability then proceeds in three steps. First, given the dose seen by the hardware, tasks are examined to determine feasibility of execution. If a task cannot be executed for lack of operational equipment, some alternative "family" of one or

more tasks is substituted. Second, once this task sequence is established, the dose seen by man is considered; and the task performance parameters (average and standard deviations of task times and the probabilities of task success) are adjusted to reflect the expected degradation or radiation-induced performance decrement. Third, these adjusted input data are then used with the Siegel-Wolf model to determine whether the task sequence can be completed within the time allowed by the mission constraints.

If the first step is eliminated, one can explore the impact of hardening the hardware. If the second step is eliminated, one can assess the impact of inflight malfunctions or conventional weapon battle damage. If one substituted a yet-to-be-developed chemical bacteriological warfare (CBW) performance decrement routine for the nuclear decrement routine developed in prior efforts, one could explore the impact of CBW environments on manned systems. In short, the modularity of the software developed permits future adaptation to other problem areas, a desirable, if not necessary, software design concept.

Studies in Progress

Two other contractual efforts are now underway. One extends the two-man model so multiple crews can be simulated in multi-phased missions. This effort focuses on the B-52 but is potentially applicable to crew station design evaluation for the B-1. Initial modelling concepts and flow charts have been prepared and programming (again, in FORTRAN IV) has been initiated.

Air surveillance and command/control systems do not have definable end points on a time continuum as do weapons delivery systems. The information processing and decision-making tasks which the crew performs also possess characteristics unlike the tasks performed in a weapons delivery system. Finally, crew structure, the redundancy in crew functions (several operators assigned to the same types of tasks--sharing the group workload), and the peacetime/wartime dichotomy of the mission profile also contrast with weapons delivery systems.

Because of these differences, AMRL is exploring several modelling concepts for treating air surveillance and command/control system simulation. Efforts focus on two objectives: (a) definition of performance requirements of man and machine, including how these requirements affect man-machine relationships; and (b) identification of human and team performance data which directly support model development. The Systems Effectiveness Branch of AMRL expects this modelling effort will identify implicit research requirements. It is also expected that other in-house empirical studies will impact the evolution of this class of multi-man team performance models.

In-house Efforts

The two-man model delivered by the contractor has been extensively revised. A set of graphics routines has been added to permit interactive interpretation of the radiation decrement data (Seifert, 1972). Output routines are also being developed to graphically portray both the raw results from simulation runs and statistical curve fitting analyses of output data.

Preliminary attempts have also been made to treat simplistically the disruptive influences of temporary flashblindness and vomiting episodes, where in each case the disruption is treated as an additional task the operator is obliged to perform. This effort is directed toward identifying the requirements for a more refined approach for modelling such disruptive effects of nuclear environments.

Rather extensive sensitivity tests are also being performed to determine how much the model's output may be affected by inaccuracy in the input data. It appears that statistical validity in a model of this sort may be obtained only at the expense of construct validity (and vice versa) if the accuracy requirements become quite critical. However, construct validity is essential if models are to aid designers. Optimal trade-offs between construct and statistical validity are yet to be determined.

Future Efforts

The applicability of GASP and P-GERTS simulation languages is being explored, since these represent state-of-the-art systems engineering methods. The intent is to modify these languages to include constructs which permit the engineer to model man as a viable element of a system. This will facilitate the incorporation of human factors considerations in models of man-machine systems by designing human engineering capabilities into advanced systems engineering methodology.

The Systems Effectiveness Branch of AMRL, in conjunction with the Weapons Effects Branch of SAM, is exploring the feasibility of somehow mimicking the radiation illness syndrome. This would permit empirical investigation of the performance degradation due to induced degradation emulating the impact a radiation environment might have on man. This would supplement the primate radiation data and provide estimates of how man's degradation in the information processing/decision-making tasks of air surveillance and command/control systems might impact the vulnerability/survivability of those systems.

Acknowledgment

The research reported in this paper was conducted by personnel of the Aerospace Medical Research Laboratory, Aerospace Medical Division, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio. This paper has been identified by the Aerospace Medical Research Laboratory as AMRL-TR-72-30. Further reproduction is authorized to satisfy needs of the U. S. government.

References

- Chubb, G. P. Using Monte Carlo simulation to assess the impact radiation-induced performance degradation can have on mission success. *Proceedings, Symposium on Computer Simulation as Related to Manpower and Personnel Planning*, Naval Personnel Research and Development Laboratory, Washington, D. C. July 1971.
- Seifert, D. J. Graphic routines for analyzing input and output in a Monte Carlo simulation. *Proceedings, Third Annual Pittsburgh Conference on Modelling and Simulation*, University of Pittsburgh, April 1972. AMRL-TR-72-12.
- Siegel, A. I., & Wolf, J. J. *Man-machine simulation models*. New York: Wiley-Interscience, 1969.
- Siegel, A. I., Wolf, J. J., Fischl, M. A., Meihle, W., & Chugg, G. P. *Modification of the Siegel-Wolf operator simulation model for on-line experimentation*. AMRL-TR-71-60.

DETERMINANTS OF THE POST-AROUSAL PERFORMANCE DECREMENT:

IMPLICATIONS FOR RESEARCH AND APPLIED PSYCHOLOGY

Robert B. Tebbs

United States Air Force Academy

The results of two experiments are presented in which post-arousal motor or perceptual-cognitive performance was measured. The combined results of the two experiments clearly demonstrated that the degree of the post-arousal performance decrement is due more to the chronic level of the S's level of anxiety than to the time of night from which the awakening was made, or to the REM or NREM sleep stage preceding awakening.

How well can a crew member perform after abrupt arousal from sleep? The data from past research provides overwhelming evidence that performance in the post-arousal period will be approximately 25% less efficient than it will be during the normal waking period (Tebbs, 1971). Yet, in most studies which have measured post-arousal performance (PAP), there have been wide individual differences in the PAP decrement (e.g., see Seminara & Shavelson, 1969). Nevertheless, in some military operations commanders may be forced to awaken crews and expect them to perform efficiently in the 15-20 minute period immediately following arousal. If a commander could determine, with some confidence, those crew members who perform more efficiently than others in the immediate post-arousal period, he could task them to perform the critical PAP tasks and thus increase the probability of mission success.

The studies which will be reviewed in this paper were designed to determine which of three independent variables, i.e., time of night, the sleep stage preceding awakening, or the level of anxiety as measured by personality tests had the most effect on the PAP. In these experiments, the anxiety factor was clearly demonstrated to be the most important variable. However, before the results of those experiments are reported, it seems appropriate to briefly review the reasons why these three variables were selected for manipulation in these experiments.

When PAP has been measured during the normal nocturnal sleep period, the evidence shows, in general, that performance declines from the early part of the night and reaches its nadir about 3-4:00 A.M. Then performance improves, forming a U-shaped performance function. There have been several demonstrations of a close covariation between body temperature and the performance curve (e.g., see Kleitman, 1963). Trumbull (1966) has stated that most performance

curves can be brought into line with the cyclical variation in body temperature. In spite of the general trend, however, Hartman, Langdon & McKenzie (1965) were not able to find any relationship between body temperature and PAP measures. This finding prompted Tebbs (1966) to search for additional determinants of the PAP decrement.

A prime candidate was the effect of the preceding sleep stage on PAP. It was well known in 1966 that sleep was not a unitary period of physiological activity. Rapid-eye-movement (REM) sleep and non-rapid-eye-movement (NREM) sleep can be identified by the electrophysiological indices (Rechtschaffen & Kales, 1968). REM and NREM sleep have quite different patterns of physiological activity. The brain is more active during REM sleep than during NREM sleep, and there is a loss of some muscle tonus during REM sleep which does not occur during NREM sleep. If the level of physiological activation, as measured by brain activity and by muscle tonus, is a predictor of PAP, then the preceding sleep stage could conceivably be an important variable in predicting the level and efficiency of PAP. On the basis of the nature of the pattern of activation, it follows that performance on cognitive tasks might be better after REM than NREM sleep, and for motor tasks the PAP might be better after NREM sleep. The activation thesis is a common one in the psychological literature. The potential effects of the sleep stages on PAP have been explicitly or implicitly predicted by Kleitman (1970), Snyder (1966), and Roffwarg, Muzio & Dement (1966).

The level of anxiety, as indicated by scores on tests such as the Minnesota Multiphasic Personality Inventory (MMPI), may also be thought of as another indicator of a level of activation. While body temperature and sleep stages may indicate fluctuations in activation from the individual's baseline, the level of anxiety may be conceptualized as a measure of the individual's general level of activation.

Of special interest to this discussion is the work of Monroe (1967) in his comparison of subjects who, by self-report, were either good or poor sleepers. Monroe found, for example, that poor sleepers showed less of a decline in their body temperature from the waking state during a night of sleep than did good sleepers. Poor sleepers also scored higher on most of the MMPI scales than did good sleepers. In general, the poor sleeper seems to maintain a higher level of activation, as measured by the physiological indices, than the good sleeper. The results of the following studies show that subjects who were almost comparable to Monroe's good sleepers on MMPI scales performed better than subjects who were also nearly comparable to Monroe's poor sleepers on the clinical MMPI scales.

Strength of Grip (Tebbs & Foulkes, 1966)

Method

The independent variables in this study were time-of-night, sleep stage, and anxiety. The dependent variable was strength of grip as measured by a dynamometer. Twenty young male Ss were used. Ten were the high scorers on Byrne's Sensitization-Repression scale (1965) of the MMPI from a population of 50 Ss and 10 were the lowest scorers. The Ss were aroused 4 times on two non-consecutive nights. The REM-NREM nights were counterbalanced so that half of the Ss had the REM night awakenings first and half had the NREM awakenings first.

Results

When comparisons were made between PAP of the sensitizers (MMPI profile above the mean) and repressors (MMPI profile near the mean) no significant differences were found between their mean strength of grips. Also, the REM-NREM comparisons produced no significant differences. However, when the pre-sleep strength of grips were compared with post-awakening trials, the decrement in the strength of grip was significant for the sensitizers (Ss similar on MMPI scales to Monroe's poor sleepers), i.e., $p < .01$; but the pre-sleep to post-arousal decrement was not significant for the repressors (Ss comparable to Monroe's good sleepers on the MMPI).

Perceptual-cognitive Tasks (Tebbs, 1971)

Method

The Ss were 32 Air Force Academy Preparatory School cadets. Sixteen of the Ss were among the highest scorers (tense) on the calm-tense scale of the 16PF, and 16 were among the lowest scorers (calm) of the 190 cadets who were tested. In addition, the Ss were found to be significantly different on most of the MMPI scales (as were the strength of grip Ss and Monroe's good-poor sleepers) and on the good-poor sleep dimension as measured by a sleep questionnaire similar to Monroe's. They were not different, however, on the basis of Otis IQ test scores or on their mean scores on the College Entrance Examination Board (CEEB) test scores.

The Ss slept for two non-consecutive nights in the laboratory. The independent variables were counterbalanced. On each trial, Ss were aroused and, after being awake for about two minutes, required to perform two consecutive cognitive tasks (Visualization Tasks 1 and 2 from French, Ekstrom & Price, 1963). The first test was performed approximately 3 to 6 minutes after awakening and the second 8 to 15 minutes after awakening. Ss took about 1-2 minutes to read the instructions for each test. Thus, the testing took place in a relatively extended period following awakening.

A comparison group of 10 (tense) and 14 (calm) Air Force Academy cadets was formed to provide a means of comparing waking performance with the PAP of the experimental Ss. Scheduling problems and the limitation of having only two forms for each test were the pragmatic reasons for not using the experimental Ss as their own comparison group.

Results

The results demonstrated a practice effect. Late awakenings produced better performance than early awakenings ($p < .01$), and performance on Night 2 was better than performance on Night 1 ($p < .05$). No differences in the PAP on any of the comparisons were found on the second task. Apparently the time from awakening and the work on the first task were sufficient to bring about performance comparable to waking performance. For the first task, none of the between-trial PAP comparisons produced significant differences between the calm and tense Night Ss.

The finding of interest was the comparison of second trial of the Day comparison groups with the Night experimental Ss' fourth and final trial on the second night. The results demonstrated that the PAP of the calm Night Ss was not different from either the calm or the tense Day Ss. However, even after four trials, the PAP of the Night tense Ss was significantly different ($p < .05$) from both of the two Day comparison groups. Aptitude for the test was ruled out as an explanation for the difference as there were no significant differences in IQ scores between the Night calm and tense groups. Also, even though the Day calm Ss had significantly higher mean CEEB scores than the other three groups of Ss, there was no significant difference between their scores and the Day tense Ss or the Night calm Ss. Therefore, although the task was different in nature from the strength of grip study, and even though the time period following awakening was longer than that in the strength of grip study, the results were the same, i.e., both studies demonstrated that when the PAP was compared with waking performance it was significantly poorer for Ss whose mean MMPI profiles were above average. Ss whose mean MMPI profiles were near the mean did not have a significant waking vs PAP decrement.

Discussion and Implications

These results demonstrate that PAP is not significantly affected by the preceding sleep stage. The results do not contradict findings such as Scott's (1969) where trends for differences based on the preceding sleep stage have been noted in the period immediately following awakening (i.e., 1 to 2 minutes). The conclusion applies to the moderately extended period following awakening where the decrement appears to be more from the lingering effects of sleep per se than to the sleep stage.

The implications for military operations seem clear. Crew members who may be expected to perform mission-crucial tasks following abrupt arousal should be screened for their chronic level of anxiety. Undoubtedly more research is required to develop a suitable regression equation for the selection procedure. The common test used to separate Ss in Monroe's study and the two experiments briefly reported here was a difference based on the mean MMPI profiles, and would seem to be the test from which to start in developing the validation and selection research. While defining anxiety in terms of average or higher than average MMPI profiles may seem to be too vague, it seems to be adequate from the standpoint of validation criterion which is important to many Air Force and military operations; i.e., performance following abrupt arousal. This criterion probably has its most pertinent application to operational crews; however, it should also be considered and tested for applicability for commanders and operations officers who must sometimes make crucial decisions in the hazy period between sleep and full wakefulness.

References

- Byrne, D., Golightly, C., & Sheffield, I. The repression-sensitization scale as a measure of adjustment. *Journal of Consulting Psychology*, 1964, 29, 586-589.
- French, J. W., Ekstrom, R. B., & Price, L. A. *Manual for kit of reference tests for cognitive factors* (Rev. 1963). Princeton: Educational Testing Service, 1963.
- Hartman, B. O., Langdon, D. E., & McKenzie, R. E. *A third study on performance upon sudden awakening*. United States Air Force School of Aerospace Medicine Technical Report 65-63, August, 1965.
- Kleitman, N. *Sleep and wakefulness* (Rev. 1963). Chicago: University of Chicago Press, 1963.
- Kleitman, N. Implications for organization of activities. In E. Hartmann (Ed.), *International psychiatric clinics* Vol. 7, No. 2, *Sleep and dreaming*. Boston: Little, Brown, 1970.
- Monroe, L. M. Psychological and physiological differences between good and poor sleepers. *Journal of Abnormal Psychology*, 1967, 72(3), 255-264.
- Rechtschaffen, A., & Kales, (Eds.) *A Manual of standardized terminology, techniques and scoring systems for sleep stages of human subjects*. Public Health Service, U. S. Government Printing Office, Washington, D. C., 1968.

- 40
- Roffwarg, H. P., Muzio, J. N., & Dement, W. C. Ontogenetic development of the human sleep-dream cycle. *Science*, 1966, 152, 604-619.
- Scott, J. Performance after abrupt arousal from sleep: Comparison of a simple motor, a visual perceptual, and a cognitive task. *Proceedings of the 77th Annual Convention of the American Psychological Association*, 1969, 4, 225-226.
- Seminara, J. L., & Shavelson, R. J. Effectiveness of space crew performance subsequent to sudden sleep arousal. *Aerospace Medicine*, 1969, 40(7), 723-727.
- Snyder, F. Toward an evolutionary theory of dreaming. *American Journal of Psychiatry*, 1966, 123(2), 125-140.
- Tebbs, R. B., & Foulkes, D. Strength of grip following different stages of sleep. *Perceptual and Motor Skills*, 1966, 23, 827-834.
- Tebbs, R. B. Post-awakening visualization performance as a function of anxiety level, REM and NREM sleep, and time of night. Unpublished doctoral dissertation, University of Wyoming, 1971.
- Trumbull, R. Diurnal cycles and work-rest scheduling in unusual environments. *Human Factors*, 1966, 8(5), 385-398.

PERSONALITY CORRELATES OF LEFT-HANDEDNESS

Mark L. Stein

United States Naval Hospital, St. Albans, New York

In an investigation of the personality correlates of left-handedness, 20 left-handed and 20 right-handed male subjects were administered a battery of psychological tests designed to measure self-concept, characterological oppositionality ("contrariness"), and tendency to compensate for physical defect. Results indicated that left-handed subjects tended to be more oppositional, tended less to compensation for physical defect, and, on one variable, tended to have a different self-concept from that of right-handed subjects. These findings were interpreted as suggesting the existence of a distinctive left-handed personality style.

Left-handed individuals comprise, with minor local variations, somewhat less than ten percent of the population--roughly 20 million Americans at present.

Despite the fact that left-handedness is a statistical rarity--or perhaps, in a subtle fashion, because of this fact--the left-handed individual has throughout history been the focus of much attention, both scientific and nonscientific; attention which typically assumed the form of attribution of deviant personality characteristics to left-handed individuals.

In previous eras, the left-hander was regarded as "sinister"--the technical term for left-handedness is, in fact, sinistrality. He was regarded as congenitally criminal; as an omen of evil tidings; as a person in league with the devil; generally speaking, as a threat to the social group.

More recently, scientific attention has been directed toward the issue of whether real personality differences exist between left- and right-handers. The establishment of such differences, of course, would shed light upon the underpinnings of traditional sinistrophobic superstition.

The notion that left-handed individuals display consistent personality differences from right-handed individuals is based upon the theoretical relationship between the individual and his physical/social environment as posited by psychoanalytic ego psychology among other orientations. As ego psychology is primarily concerned with the modes by which an individual adapts to his environment through the mechanisms of personality (Hartmann, 1958), it is useful to consider

what sort of environment confronts the left-hander.

As has often been noted, the left-hander lives in a right-handed world. A large proportion of the tools and manipulanda necessary to efficient functioning in a complex society--pens, school desks, pay telephones, cooking utensils, power tools, clothing, rifles, control panels, etc.--are oriented to the 90% right-handed majority. While the social opprobrium previously associated with left-handedness has largely receded, the left-hander is still confronted with challenges ranging from mastery of the right-handed handshake to difficulty in acquiring, borrowing, or sharing athletic equipment. Such issues as sense of mastery (Hendrick, 1942) and sense of competence (White, 1963) are thus particularly apposite.

While previous systematic research into the issue has been rather sparse, data have been presented indicating an association between left-handedness and self-concept (Palmer, 1963), and between left-handedness and "oppositionality," a concept which refers to characterological contrariness or negativism (Blau, 1946; Finn & Neuringer, 1968). Deriving from the above, the presently-reported program of research has attempted to establish the existence of a left-handed personality *style*, a construct consisting of several interrelated personality variables hypothesized to be associated with left-handedness.

Hypotheses

1. Oppositionality: Left-handed individuals were hypothesized to be significantly higher in oppositionality than right-handers.
2. Self-concept: Left-handed individuals were hypothesized to differ in self-concept from right-handers in terms of: sense of autonomy, self-confidence, motor awkwardness, social awkwardness, self-regard.
3. Compensation for physical defect: On the basis of the fact that left-handers are uniquely required by the environment to acquire a degree of skill with the nonpreferred hand, left-handers were hypothesized to demonstrate a significantly higher tendency than right-handers to compensate for, or strive to overcome, perceived physical defect.

Method

Subjects

Forty male hospital corpsmen volunteered as subjects, twenty left-handed (LHed) and twenty right-handed (RHed). The two groups did not differ significantly in age or intelligence.

Apparatus

The three above hypotheses were defined operationally in terms of psychological tests.

1. Oppositionality was measured in terms of three test scores: (a) amount of white-space perceived on the Rorschach test (white-space perception, as it involves a figure-ground reversal in contravention of test instructions, is often interpreted as a "sign" of oppositionality), (b) oppositionality ratings made by each S's immediate supervisor on a semantic-differential type seven-point scale with big higher values indicating greater rated oppositionality, and (c) score on a story-completion scale measuring the extent to which S admits to being oppositional.

2. Self-concept was measured in terms of five relevant scales derived from the Adjective Check List (Gough, 1965).

3. Tendency to compensate for physical defect was measured in terms of scores on a story-completion task derived from TAT-type stimuli with thematic material involving solutions to problems posed by physical defect.

Results

A summary of total test scores and statements of significance is presented in Table 1.

It is seen in Table 1 that on two of the three Oppositionality measures, LHers differed from RHers in the predicted direction. LHers were significantly higher than RHers in terms of white-space scores. As LHers were also significantly more productive on the Rorschach, producing almost twice as many percepts as RHers, the space-score for each subject was adjusted by dividing it by his response total. It is seen that LHer's adjusted space-score remained significantly higher than those of RHers.

On the Oppositionality Rating measure, LHers were rated as significantly more oppositional than RHers. The differences between LHers and RHers on the Control Rating items were not significant, indicating that the raters were able to discriminate between rating-variables. Additional confirmation of the theoretical association between white-space and oppositionality was provided by the rank-order correlations between space-score and oppositionality ratings, which were .595 and .590 for LHers and RHers, respectively, both correlations significant at the .01 level.

The Story Completion task assessing "willingness to admit to being oppositional" yielded no significant differences between handedness groups.

Table 1
Summary of Results

Test-scores	Left-handed	Right-handed
Rorschach S ^a	96.000	18.000* ^b
S/Number Responses	2.779	1.134**
Oppositionality Ratings	279.000	182.000***
Oppositionality Control Ratings	231.000	230.000
Oppositionality Story Completion	204.000	223.000
ACL Autonomy	5.050	2.700****
ACL Self-confidence	6.400	5.900
ACL Motor-Awkwardness	0.900	0.150
ACL Social-Awkwardness	0.750	-0.750
ACL Low Self-Regard	-2.700	-3.800
Compensation Scale	190.000	236.000*****

^aS = White-space score: scoring based upon two points for each "pure" space response and one point for each space response that included solid portions of blot.

^b = All significance statements refer to differences between all LHers and all RHers for the test in question.

* F = 33.0, $p > .01$, $df = 1, 36$
 ** F = 8.36 $p > .01$, $df = 1, 36$
 *** F = 15.1 $p > .01$, $df = 1, 36$
 **** F = 1.74 $p > .05$, $df = 38$
 ***** F = 7.40 $p > .05$, $df = 1, 36$

Of the five self-concept scales, only the Autonomy scale yielded a significant difference, with LHers describing themselves as significantly more autonomous than RHers. While there was a tendency for LHers to describe themselves as more awkward, motorically and socially, and as of lower self-esteem than RHers, the differences did not attain significance. Surprisingly, LHers described themselves as more self-confident than RHers, but again the difference was not significant.

On the story-completion items assessing the subject's tendency to overcome or compensate for physical defect, RH subjects scored significantly higher than did LHers. There were no significant differences on the control items.

Discussion

Preliminary statistical analysis of the data strongly suggests the existence of differences between LH subjects and RH subjects on a number of personality variables.

As the most striking differences occurred in the area of oppositionality, it is useful to consider the theoretical substratum from which the hypothesis derived. At least two possibilities have been advanced. Blau (1946), in a highly controversial monograph, suggested both that left-handedness is purely a social phenomenon and that left-handed individuals become left-handed through an early acquired or congenital tendency to negativism or contrariness. Left-handedness, for Blau, constituted merely one expression of this negativism. Blau's notions, however, have been largely discredited by the firm establishment of the hereditary transmission of left-handedness. The focus of attention may thus be profitably shifted to an interpretation of oppositionality as an adaptive maneuver developed as one characterological repercussion of left-handedness, rather than as a cause.

What is the adaptive significance of this tendency of left-handers to swim upstream against the social current? Viewing the situation in ego psychological terms, the motoric and social burdens imposed by left-handedness may be supposed to place the young left-handed individual in a weaker, less effective position than his right-handed counterpart in relation to a complex, demanding environment. To use a psychoanalytic term metaphorically, the left-hander, as a result of such early experience, may develop a sort of "character armor," (Reich, 1967); in more contemporary parlance, we may speak of a character *style* affording the left-hander a measure of imperviousness against disturbing environmental stimuli. In other words, the issue may be viewed in terms of autonomy, the left-hander developing an oppositional style as a means of establishing his autonomy in the face of an encroaching, intractable environment.

Several strong arguments support interpretation of the data predominantly in terms of autonomy-strivings. The first argument rests upon the fact that on the self-concept measure, left-handers scored significantly higher on Autonomy than did right-handers, suggesting that autonomy is a more salient issue for the former group. Second, a body of previous research literature exists that supports the relationship between white-space on the Rorschach and autonomy-strivings, suggesting an intimate linkage between the concepts of oppositionality and autonomy-striving (reviewed by Fonda, 1960).

The finding that right-handers tend to compensate for physical defect to a significantly greater degree than left-handers was contrary to prediction. It is possible that left-handers tend to channel compensatory strivings into nonphysical modalities, for example through heightened achievement motivation. This will be one fruitful focus for future research.

Analysis of the data, however, has not yet been completed. The task remains to run the data through a multiple correlation computer program to determine which of the variables, or sets of variables, have the greater predictive weight in discriminating between left-handed and right-handed individuals. The ultimate objective of the research is to isolate a distinct left-handed personality style (Shapiro, 1965), a configuration of personality variables statistically powerful enough to discriminate blindly between left-handed and right-handed individuals in random population samples.

While this research presently rests within the sphere of personality theory, it is anticipated that it will ultimately have a number of practical applications. From a personnel psychology standpoint, the establishment of a distinct left-handed personality style is expected to be of assistance in large scale screening programs. Merely through indicating that he is left-handed, a subject will provide a wealth of personality data useful to supervisors, personnel officers, etc. Further, an appreciation of the implications of left-handedness may be of use in training programs, with special attention given to left-handed students on relevant tasks, consistent with their unique strengths and weaknesses.

Finally, data regarding left-handed personality styles should be of interest and assistance to workers in clinical areas, in terms of research, diagnostic personality assessment, and treatment.

References

- Blau, A. *The master hand*. New York: American Orthopsychiatric Association, 1946.
- Finn, J., & Neuringer, C. Left-handedness - A study of its relation to oppositionality. *Journal of Projective Techniques*, 1968, 32, 49-52.
- Fonda, C. The white-space response. In Rickers-Ovsiankina (Eds.) *Rorschach psychology*. New York: Wiley, 1960.
- Gough, H. *The adjective check list manual*. Palo Alto: Consulting Psychologists Press, 1965.
- Hartmann, H. *Ego psychology and the problem of adaptation*. New York: International Universities Press, 1958.
- Hendrick, I. Instinct and the ego during infancy. *Psychoanalytic Quarterly*, 1942, 11, 33-58.
- Palmer, R. Hand differentiation and psychological functioning. *Journal of Personality*, 1963, 31, 445-461.
- Reich, W. *Character analysis*. New York: Essandess, 1967.
- Shapiro, D. *Neurotic styles*. New York: Basic Books, 1965.
- White, R. *Ego and reality in psychoanalytic theory*. New York: International Universities Press, 1963.

All-Volunteer Force

CHAIRMAN: Captain John M. Bermudez

PREDICTION OF DISENROLLMENT AND MILITARY APTITUDE AT
THE NAVAL ACADEMY WITH THE STRONG
VOCATIONAL INTEREST BLANK

Norman M. Abrahams and Idell Neumann

Naval Personnel & Training Research Laboratory

This research examined the potential of the Strong Vocational Interest Blank for improving selection of Naval Academy midshipmen. Using data of two classes, empirical scales were constructed to predict academic and motivational disenrollment and military aptitude ratings. All three scales yielded significant relationships with their respective criteria in cross-validation samples. The data clearly support the validity of the Strong Vocational Interest Blank in predicting various criteria of success at the Academy.

Despite rigorous selection requirements, only 65% of those admitted to the U. S. Naval Academy graduate as commissioned officers. Since the Academy is a major as well as expensive source of Navy officers, those selected should have the highest likelihood of success as students and officers.

During the last few years research has been devoted to evaluating the Strong Vocational Interest Blank (SVIB) as a means of forecasting success at the Academy. This report describes research on the validity of the SVIB for the prediction of disenrollment and military aptitude. Approximately one-third of each entering class ultimately disenrolls for motivational, academic, medical, or other reasons. Military aptitude, which is based on ratings of attitude, leadership, officer potential, performance of duty, and bearing and dress is related to subsequent performance as an officer.

Currently, the Academy's primary screening device is a weighted composite, referred to as the candidate multiple, that includes four cognitive test scores, high school standing, high school activities, and recommendations. This composite and its components have been found to predict academic attrition and military aptitude but not voluntary motivational disenrollment (Howland, 1971). This lack of validity provided the original impetus for research with the SVIB.

Procedure

Subjects and instrument

Members of the Naval Academy classes of 1971-1973 completed the 1966 edition of the SVIB within their first week at the Academy. Applicants to the class of 1974 completed the SVIB on a voluntary basis as part of the application procedure. In the fifth year of testing, for selection of the class of 1975, the SVIB became a required part of the application packet.

Disenrollment analyses

Initial analysis indicated that voluntary motivational and academic disenrollments account for 53% and 30%, respectively, of all disenrollments. Since the remaining 17% are divided among medical, conduct, aptitude, etc., empirical scales were constructed only for academic and motivational disenrollment.

The ideal procedure for scale construction would have been a comparison of the SVIB responses of members of each of the two disenrollment categories with those of graduates. However, as of March 1970, when the disenrollment information was obtained, none of the classes for which SVIB data were available had yet been commissioned. Thus it became necessary to use only the remaining midshipmen of the 1971 class, since they had the greatest likelihood of receiving their commissions.

For scale construction purposes, the remaining class of 1971 was divided into a key-development and cross-validation sample containing 70% and 30% of the 934 cases, respectively. The academic and motivational disenrollees of the 1971 and 1972 classes were each similarly divided into key-development and cross-validation samples. The response proportions of the remaining midshipmen were compared with the response proportions of each disenrollment group. All item responses with a percent difference of 10 or greater were included in the academic and motivational disenrollment keys. Each key was then cross-validated on its appropriate disenrollment criterion, as well as on the other category of disenrollment.

Military Aptitude Analyses

The second criterion, military aptitude rating, is assigned at the end of each semester and has been found to be extremely stable across semesters. Consequently, to insure comparability on the rating criterion, first-year ratings for the 1971-1973 classes were used.

The procedures employed in predicting military aptitude sought to measure criterion variance that is not currently predicted. For this reason it became necessary to apportion each midshipman's aptitude

rating into a currently predictable and unpredictable portion. Using the candidate multiple in a regression equation based on the 1971 and 1972 classes, we computed each midshipman's predicted aptitude rating. The predicted aptitude rating was subtracted from his actual aptitude rating, thereby leaving the currently "unpredictable" or residual portion of the criterion against which to validate SVIB items and scales.

Next, all midshipmen of the 1971-1973 classes were scored on the recently developed "basic scales" of the SVIB (Campbell, 1971). Each of these scales measures interest in one type of activity or in closely related activities; thus, they cover homogeneous clusters of interests represented in the SVIB, such as science, mechanical, social service, medical, etc. Correlations between the 22 basic scales and the residual aptitude criterion were computed for each class and for the class of 1972. A multiple-regression analysis was also conducted.

Since an empirically-derived key might exceed the validity of the basic scales, an item analysis of the SVIB was conducted using the residual criterion scores of the class of 1971. The SVIB items that correlated with the criterion were identified and unit-weighted in accordance with Campbell's (1971) dimensionality procedure for cross-validation on the class of 1972.

Finally, the items found valid were rationally clustered in an attempt to increase understanding of the predictive factors as well as to increase validity. It was reasoned that differential weighting of item clusters might prove to be more valid and reveal possible inadequacies in the item pool.

Results and Discussion

Prediction of Disenrollment

Academic disenrollment. The item analysis conducted for differentiating academic disenrollees from remaining midshipmen provided 56 item responses with a percentage difference of 10 or more. This scoring key, applied back to the key-development sample, yielded a biserial correlation of .55, which reduced to .24 on cross-validation. Despite the considerable shrinkage, this validity provides a significant relationship between the scale and academic attrition.

When motivational disenrollees are scored on this scale, their mean is intermediate between academic disenrollees and remaining midshipmen, indicating that drop categories are not entirely independent.

Motivational disenrollment. The motivational disenrollment scale, constructed by comparing the item responses of remaining midshipmen and motivational disenrollees, contains 121 item responses. This scale

correlated .72 with motivational attrition in the key-development sample and .36 in the cross-validation sample. When the cross-validation disenrollees were divided into early and late motivational drops the validities, respectively, were .42 and .31. This difference, together with a comparison of mean scores, suggests an underlying continuum of motivation on which the early drops are lowest, later drops intermediate, and remaining midshipmen highest.

The mean score of academic drops on this scale is intermediate between motivational drops and remaining midshipmen. Again, this indicates a lack of independence in these disenrollment categories.

Table 1, an expectancy chart based on the motivational disenrollment scale, shows a consistent relationship with the probability of motivational disenrollment. Those scoring in the bottom fifth are over three times as likely to disenroll as those in the top fifth. The scale is also valuable for identifying other drops, since 21% of those in the bottom fifth versus 10% in the top fifth disenroll for all other reasons.

Although these findings show excellent potential for predicting disenrollment from vocational interests, a similar study (Spense, Sena, & Westin, 1971) among Air Force Academy cadets yielded negative results. Three groups of cadets were compared on 19 scales of the Kuder Form DD judged relevant to the Academy. One group consisted of cadets who disenrolled because of changes in career goals or problems in adjusting to a military environment; the second group consisted of cadets disenrolled for all other reasons; the third group consisted of the remaining cadets. The authors concluded that there was little hope of discriminating between unsuccessful and successful Academy cadets in terms of interests as measured by the Kuder Occupational Survey. However, since they only used the standard scales and did not attempt the development of an empirical scale, it is premature to conclude that the Kuder is incapable of predicting Air Force cadet success.

Prediction of Military Aptitude

Basic scales. Of the 22 SVIB basic scales, only three were significantly related to residual aptitude scores across all three classes. Two scales, Recreational Leadership and Agriculture, correlated positively while the Music scale correlated negatively with the criterion. The most consistently valid scale, with positive correlations ranging from .17 to .19, was Recreational Leadership. It contains, for the most part, items reflecting sports and athletic interests.

For the class of 1972, the SVIB basic scales were entered into multiple-regression analysis. The four scales that added significantly provided a multiple R of .227 with an estimated cross-validity of .207.

TABLE 1

Expectancy Table for SVIB Motivational Disenrollment Scale
Based on 1971-1972 Naval Academy Midshipmen

SVIB Score Category	Expected Number in Class	Expected Number and Odds in each Criterion Category					
		Graduates		Motivational Disenrollees		All Other Disenrollees	
		N	Odds	N	Odds	N	Odds
HIGHEST 18% (≥ 139)	239	192	80 in 100	24	10 in 100	23	10 in 100
NEXT 20% (133-138)	272	192	70 in 100	32	12 in 100	48	18 in 100
MID 21% (128-132)	277	182	66 in 100	47	17 in 100	48	17 in 100
NEXT 21% (120-127)	289	172	60 in 100	62	21 in 100	55	19 in 100
LOWEST 20% (≤ 119)	271	126	47 in 100	88	32 in 100	57	21 in 100

This multiple R is based on the following scales presented in the order in which they entered the regression: (a) Recreational Leadership, (b) Teaching, (c) Mathematics, and (d) Technical Supervision.

Empirical scale. In the 1971 class, which was used for item analysis, the empirical SVIB scale correlated .34 with the residual aptitude criterion. The first cross-validation, on the class of 1972, yielded a validity of .28 against the residual criterion, while the second cross-validation, using the 1973 class, provided a validity of .24. These validities exceed the basic scale composite and can be regarded as incremental since the present selection composite is, by design, not correlated with the residual scores.

The validities of the empirical aptitude scale against actual aptitude ratings, in contrast to residual aptitude ratings, were .31 and .26 for the 1972 and 1973 classes, respectively. These are virtually identical to the validities of .31 and .27 for the candidate multiple on the same samples. A linear combination of the SVIB scale and the candidate multiple almost doubles the proportion of criterion variance accounted for, with multiple correlations of .41 and .36 for the 1972 and 1973 classes, respectively.

Briefly, concerning the prediction of aptitude, it appears that the SVIB scale (a) has about the same level of validity as the existing composite, (b) is virtually independent, statistically, of the existing composite, and (c) provides a significant and practical improvement in predicting aptitude.

Examination of the items in this scale resulted in nine clusters of items, each relatively homogeneous in content. To determine whether differential weighting of the clusters would provide a higher validity than the empirical unit-weighted scale, nine separate scores were entered into multiple regression for the 1972 class. The multiple correlation itself, without correcting for shrinkage, failed to exceed the validity of the empirical scale. Thus, while the rational appeal of differentially weighting item clusters remains, the empirical results are negative.

Conclusions

Vocational interests as measured by the SVIB are significantly related to Naval Academy success, as assessed by disenrollment and military aptitude.

The empirically-developed scales provide significant increments to the current levels of predictive accuracy, and also exceed the validity of the basic scales in predicting military aptitude.

Early motivational disenrollment is more predictable than later motivational disenrollment. This finding suggests that the early drops may be lowest on a continuum of motivation.

Motivational disenrollment is more predictable from SVIB responses than is academic disenrollment. Among other factors, this could be due to more stringent selection on academic ability, thereby restricting the range or to the nature of the instrument itself. Since previous studies with the SVIB have shown higher relationships for predicting academic criteria the former interpretation seems more plausible.

Further exploration into identifying, expanding, and refining the dimensions represented in the empirical scales is likely to further improve prediction.

References

- Campbell, D. P. *Handbook for the Strong Vocational Interest Blank*. Stanford University Press, Stanford, California, 1971.
- Howland, R. W. *Report of recruitment research study group*. Memorandum for the Superintendent, U. S. Naval Academy, Annapolis, Maryland, July 1971.

A PRELIMINARY ANALYSIS OF THE 1970 OTS APPLICANT POOL

BY DRAFT VULNERABILITY CATEGORY

William E. Alley

Air Force Human Resources Laboratory

AFOQT test records of OTS applicants were examined to determine the relationship between draft-vulnerability and the size and quality of the applicant pool. Ordinal position in the 1970 random lottery sequence served as the basis for estimating vulnerability to the draft. Preliminary findings indicate (a) a significant proportion of OTS applicants were draft-induced to seek an AF commission during 1970, (b) qualitative differences between high, moderate, and low-vulnerability subgroups were slight as measured by performance on the AFOQT, and (c) the overall size of the applicant pool appears to have declined during 1966-1971 time period. This trend coupled with the possible loss of draft-induced candidates may necessitate corrective actions to maintain an adequate flow of qualified applicants into the OTS program.

The Air Force currently relies on three primary sources for the procurement of officer personnel: the U. S. Air Force Academy, the Air Force Reserve Officer Training Corps (AFROTC) and Officer Training School (OTS). There has been some speculation that the flow of male applicants into these programs is influenced at least indirectly by the military draft. As the services move toward a reduced or zero-draft environment, it becomes increasingly necessary to assess the influence of Selective Service on the size and quality of the officer applicant pool and to determine the characteristics of those men likely to seek Air Force commissions in the absence of the Draft.

Considerable research has been conducted recently on the topic of an all-volunteer military service. However, most studies have dealt exclusively with the enlisted force (Valentine & Vitola, 1970; Vitola & Valentine, 1971; Brunner, 1971; USAF Saber Volunteer Report, 1971). Little has been done to date in an effort to gauge the impact of a zero-draft environment on officer procurement. Guinn et al. (1971), in a study of AFROTC enrollments, found that a significant proportion of the 1971 cadet population was draft-induced to enter training. Moreover, cadets identified as draft-vulnerable were slightly better qualified overall than were true volunteer cadets.

The purpose of this study is to provide some preliminary findings on the relationship between draft vulnerability and the quantity and quality of applicants to OTS. The study was specifically designed to

explore the following content areas: (a) the extent of draft-inducement in the OTS applicant pool, (b) qualitative differences, if any, between applicant subgroups categorized according to their vulnerability to the 1970 draft, and (c) the flow, disposition and volunteer mix of applicants to OTS in 1970-71. The data, spanning the first year of the draft lottery system (1970), are to be considered suggestive rather than definitive. They are intended to define potential problem areas and to guide further research. Implications of findings for the procurement of an all-volunteer OTS officer force are discussed.

Method

AFOQT test record cards maintained at the Human Resources Laboratory served as the primary source documents for the study. Information was obtained from representative samples of male OTS applicants who were administered the Air Force Officer Qualifying Test (AFOQT) during CY 1970 (N = 3876). The following data were recorded for each applicant: date of birth, entry status as of September 1971, and the AFOQT Officer Quality (OQ) and Pilot composite scores. Each candidate's perceived vulnerability to the draft was estimated from the ordinal position of his birth date in the 1970 lottery sequence. It was assumed that those applicants with high lottery numbers (indicating little if any vulnerability) were representative of those persons who might have applied in the absence of the draft. Various multivariate distributions of the data were prepared to characterize the applicant pool.

Characteristics of the Data Base

Approximately 25,000 test records were considered in all. Table 1 provides estimates of the overall size and configuration of the 1970 applicant pool. Potentially qualified applicants were defined as those who met or exceeded the minimum standards for Air Force Commissioning (i.e., attained a score equivalent to the 25th percentile on the AFOQT-OQ composite). Samples from the two primary interest groups underlined in Table 1 were extracted and analyzed. Included were records of the entire male-entrant cohort (N = 2,806) and a sample of male non-entrants representing 9.1% of the estimated population (N = 1,070).

Results

Extent of Draft Inducement among OTS Applicants

To determine the effect of the draft on the male applicant pool, the 366 ordinal positions in the 1970 lottery were grouped into 34 categories. The first 33 contained 11 numbers each, while the 34th contained the remaining three numbers. The frequency of applicants in each category was then tabulated. It was hypothesized that if the

Table 1

Estimated Size and Configuration of the 1970 OTS Data Base

Applicant Category	Percent of Total AFOQT Test Records (N=25,106)
A. OT /Pilot/Nav Applicants	80
Total Unqualified (OQ < 25)	15
Total Qualified (OQ \geq 25)	65
Total Qualified Female	7
Total Qualified Male	58
Qualified Male Entrants	11
Qualified Male Non-entrants	47
B. Miscellaneous Applicants	20
C. Total AFOQT Test Records	100

Estimates are based on a weighted combination of entrant and non-entrant samples adjusted for sample size.

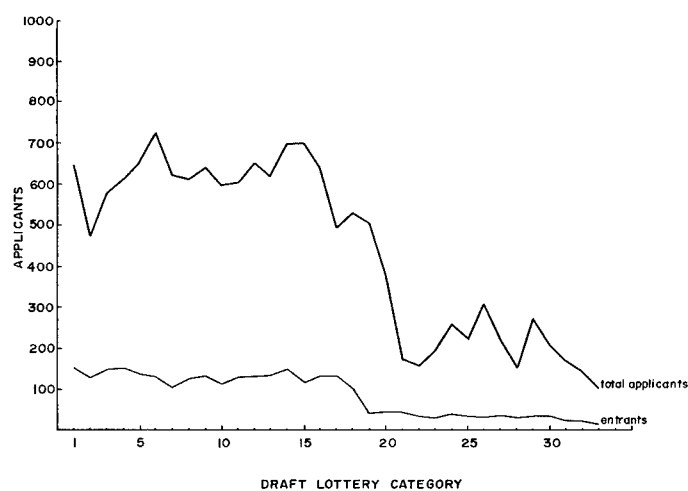


Fig. 1. Frequency distribution of qualified male OTS applicants and entrants by 1970 Draft Lottery Category.

draft was of little or no consequence to potential applicants, the number within each of the first 33 lottery categories would be roughly comparable. Figure 1 shows the distribution in graphic form for the entrant subsample and for all applicants combined. It is evident from the graph that neither of the groups is equally distributed across lottery categories. The frequencies at higher levels of draft vulnerability (lottery number 1-122) are approximately three times that found at lower levels. The second plateau beginning with categories 20 and 21 (lottery numbers 210-231) seems to represent a baseline rate; one which might be expected in a reduced or zero-draft environment.

Table 2, in which lottery numbers have been regrouped into thirds, shows that the proportion of applicants in the low-vulnerability categories (lottery numbers 245-366 was not consistent across entry status groups. In comparison with entrants, the non-entrant cohort appeared to contain a higher relative proportion of what might be considered true volunteers. Overall, the data suggest that substantial numbers of OTS applicants were draft-induced to seek commissions during 1970. Moreover, the effects of draft-inducement are evident to a differing degree in both entrant and non-entrant samples.

Table 2

Percentage Distribution of Qualified Male OTS Applicants ¹
by Entry Status and Draft Lottery Category

Lottery Sequence	Entrants % (N = 2806)	Non-entrants % (N = 1067) ²
1 - 122	52	46
123 - 244	37	38
245 - 366	11	16
Total	100	100

¹ OTS Applicants tested in CY 1970

² Excludes cases with unknown lottery sequence (N = 3)

Relationship of Applicant Quality to Volunteer Status

There has been some concern expressed over the possibility that a reduced or zero-draft environment would adversely affect the quality of Air Force accessions, especially with regard to the enlisted force. In order to consider the question of whether the quality of OTS applications would be similarly affected, information on the AFOQT performance of high, moderate, and low draft-vulnerability groups was obtained. Table 3 shows a percentage distribution of AFOQT - OQ scores for applicants categorized by entry status and draft vulnerability.

In both entrant and non-entrant samples, the low-vulnerability subgroups seem to score somewhat lower overall than do applicants categorized as highly vulnerable to the draft. The differences, however, are slight and not at all consistent. For non-entrants, the low-vulnerability group contributed proportionately fewer applicants in both the highest and lowest OQ levels.

When performance on the AFOQT - Pilot composite is compared across lottery categories (Table 4), the trend shifts in the opposite direction. It appears that the low-vulnerability subgroups in both samples score proportionately higher than do the high-vulnerability subgroups. These findings were quite unexpected, particularly in view of the earlier study of AFROTC cadets (Guinn, et al., 1972). Further research with increased sample size may provide additional clarification. In the interim, it must be concluded that quality decrements in the applicant pool, if they occur at all as a function of reduced draft pressure, do not seem to present a serious problem at present. It should be noted, however, that the question of applicant "quality" and its effect on the caliber of personnel selected for training cannot be considered independent of the overall quality of applicants. For example, a significant decrease in the size of the applicant pool may alter selection ratios to the point where the Air Force would have to accept persons of lower quality, even if the overall quality of the applicant pool remained stable.

Quantity of OTS Applicants

The question of most immediate concern to OTS is whether the yearly number of applicants anticipated in a zero-draft environment will be sufficient to maintain officer production requirements. Historically, the picture is not encouraging. Even during the Vietnam buildup when draft calls were relatively high, there is evidence of a substantial reduction in the number of persons applying to OTS. In the past six years, for example, the number of AFOQT tests given each year in support of the OTS program has decreased by more than 50 percent. Figure 2 shows this trend graphically by calendar year. In 1966, over 45,000 persons were administered the AFOQT by USAF

Table 4

Percentage Distribution of Qualified Male OTS Applicants¹ by Entry Status,
Draft Lottery Category and AFOQT Pilot Composite

AFOQT-Pilot Composite Range	Male Entrants ²			Male Non-entrants ³			
	Lottery Group		Total	Lottery Group		Total	Total
	(1-122)	(123-244)		(1-122)	(123-244)		
	N=1428)	(N=1027)	(N=2771) ⁴	(N=483)	(N=410)	(N=174)	(N=1067) ⁵
Below 25	1.9	1.6	2.0	12.2	9.0	6.3	10.0
25-35	7.0	7.5	6.9	18.8	17.8	17.8	18.3
40-50	19.4	16.2	17.4	23.8	22.4	24.1	23.3
55-65	23.5	28.5	25.1	17.8	17.6	14.4	17.2
70-80	21.2	17.3	20.9	13.1	16.1	15.5	14.6
85-95	27.0	28.9	27.7	14.3	17.1	21.9	16.6
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

¹OTS applicants tested in CY 1970

²Entrant sample is 100% of estimated population

³Non-entrant sample is 9.1% of estimated population

⁴Excludes cases with OQ < 25 (N=35)

⁵Excludes cases with unknown lottery sequence (N=3)

Table 3

Percentage Distribution of Qualified Male OTS Applicants¹ by Entry Status,
Draft Lottery Category and AFOQT-Officer Quality Composite

AFOQT-OQ Composite Range	Male Entrants ²			Male Non-entrants ³				Total	Total
	Lottery Group (1-122)	Lottery Group (123-244)	Lottery Group (245-366)	Lottery Group (1-122)	Lottery Group (123-244)	Lottery Group (245-366)			
	(N=1426)	(N=1026)	(N=316)	(N=483)	(N=410)	(N=174)	(N=1067) ⁵		
25-35	13.7	15.8	18.0	30.4	27.1	25.8	28.4		
40-50	10.2	11.1	11.1	13.7	15.1	18.4	15.0		
55-65	17.7	19.5	20.2	16.6	14.2	19.0	16.0		
70-80	22.2	18.2	14.6	12.6	16.3	14.4	14.4		
85-95	36.2	35.4	36.1	26.7	27.3	22.4	26.2		
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0		

¹OTS applicants tested in CY 1970

²Entrant sample is 100% of estimated population

³Non-entrant sample is 9.1% of estimated population

⁴Excludes cases with OQ < 25 (N=35) and cases with unknown OQ scores (N=3)

⁵Excludes cases with unknown lottery sequence (N=3)

Recruiting Service. By 1971, this figure had dwindled to slightly less than 20,000.

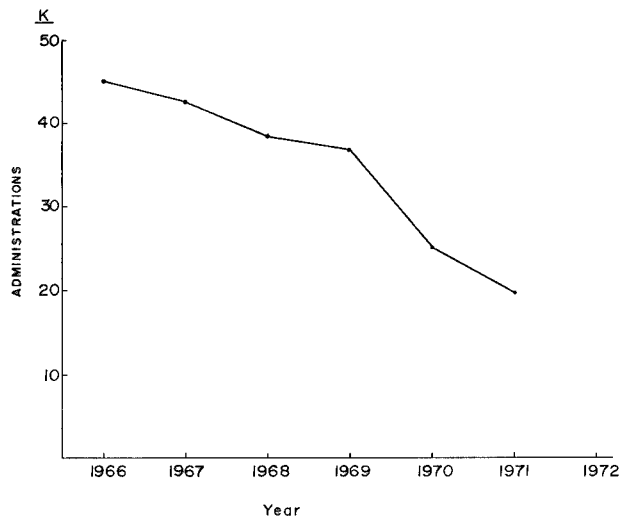


Fig. 2. AFOQT Test administrations by calendar year.

Whether this trend reflects an actual decrease in the size of the applicant "pool" has not yet been determined. Numerous other factors could have intervened during this period to make the decline more apparent than real. Between 1966 and 1970, for example, the training emphasis at OTS shifted from predominately non-rated production to predominately rated production. In addition, recruiting quotas for rated personnel were moderately curtailed in FY 71 as compared with FY 70. The influence of either of these factors could have made these data an artifact of changing AF personnel requirements. On the other hand, there is also evidence that other factors have been operating recently which did, in fact, reduce the overall availability of "draft-induced" applicants to OTS. In 1970, the draft lottery system became fully operational. In addition to providing for random selection of qualified registrants, the lottery system reduced the period of draft-vulnerability from seven years to one and modified induction procedures whereby, beginning in 1971, the youngest draft-eligibles would be selected before the oldest. To insure that draft exposure would be equitable during the transition year--1970, all males in the 19 to 26 year-old age group were assigned lottery numbers. Thus, at the end of 1970 all of the 1970 draft-eligibles, with the exception of those receiving deferments, were exempted from further vulnerability.

Inductions in subsequent years were to be drawn from a pool consisting primarily of newly eligible 19 year-olds and the residual of those men in the original lottery who had relinquished

their deferments. In this manner, Selective Service policy changes reduced the number of draft eligibles in the 1970-1971 time period. The available pool was further reduced as a result of the fact that draft calls were lower in 1971 than in 1970. The highest lottery numbers reached in both years were 195 and 125 respectively.

Further analysis may establish with greater certainty whether AFOQT test administrations are reliable indicators of the OTS applicant supply. Assuming that they do provide at least an approximation of the total number available for testing, what are the implications of current levels for OTS procurement under an all-volunteer concept? To begin with, it should be noted that the total number of AFOQTs administered annually (represented in Figure 2) is far in excess of the proportion within each year that can be considered fully qualified and willing to accept commissions. For CY 1970, the year in which the present study samples were taken, the total number tested (25,106) was composed of OTS applicants as well as smaller numbers interested in other commissioning programs (see Table 1). Estimates of the available pool must also account for the progressive reduction which takes place during the actual processing of the applications.

Annual reports outlining the disposition of male applicants to OTS for FY 70 and 71 show, for example, that of the total applications submitted (17.5K and 11.K respectively) only 12.8K were fully qualified in FY 70 and 9.2K in FY 71. Of the number fully qualified each year, even smaller proportions were willing to accept commissions if selected and invited. Declination rates for FY 70-71 varied between 32 and 38 percent. Elimination of draft inducement would probably reduce still further the number of willing and qualified applicants by a factor of at least 50 percent and possibly more.

Pending further verification of the apparent decline in the rate of OTS applications, the present findings suggest that substantial efforts may be required in the near future to maintain an adequate flow of applicants into the OTS program. These might include but not be limited to improvements in monetary and non-monetary commissioning incentives, augmentation of recruiting efforts, and the development of more efficient methods of utilizing existing applicant resources.

Conclusions

The preliminary findings of this study can be summarized as follows:

1. A significant proportion of OTS applicants were draft-induced to enter service during 1970. These findings were in accord with previous studies of both enlisted and officer (AFROTC) accessions. The relative percentages of low vulnerability applicants

appeared to be greater among non-entrants than among entrants.

2. Qualitative differences between applicant groups categorized according to draft vulnerability were slight as measured by performance on the AFOQT. Applicants within low vulnerability to the draft (true volunteers) seemed to score lower on the OQ composite than did the high-vulnerability subgroups. Their performance in the Pilot composite, however, was somewhat better than that of the draft-induced candidates. These observations held for applicants who entered training and for those who did not.

3. The number of AFOQTs administered in support of the OTS program has decreased by more than 50 percent in the 1966 to 1971 time period. If these data are indicative of a general decline in applicant availability over time, corrective policy actions may be necessary to maintain a high degree of selectivity in the program. The extent of draft-inducement among 1970 applicants suggests that the available pool of qualified and willing applicants will be further diminished with the elimination of the draft as an influence on officer procurement.

4. Additional research is necessary to verify the preliminary findings of this study and to further explore methods for increasing the size and quality of the applicant pool available for entry into OTS.

References

- Brunner, G. L. *The importance of volunteer status: An analysis and reliability test of survey data.* R-717-PR. Santa Monica, Calif.: The Rand Corporation, 1971.
- Guinn, N., Alley, W. E., & Farmer, C. B. *Impact of an all-volunteer force on AFROTC officer procurement.* AFHRL-TR-71-00. Lackland AFB, Texas: Air Force Human Resources Laboratory, 1971.
- Valentine, L. D., & Vitola, B. M. *Comparison of self-motivated Air Force enlistees with draft-motivated enlistees.* AFHRL-TR-70-26. Lackland AFB, Texas: Air Force Human Resources Laboratory, 1970.
- Vitola, B. M., & Valentine, L. D. *Characteristics of Air Force enlistees related to draft-vulnerability.* AFHRL-TR-71-29. Lackland AFB, Texas, Air Force Human Resources Laboratory, 1971.
- USAF Saber Volunteer Report. *An analysis of problems associated with the establishment of an all-volunteer force for the United States.* (Final Report). Washington, D.C.: Assistant Chief of Staff, Studies and Analysis, Headquarters, USAF, 1971.

AN ANALYSIS OF THE IMPACT OF VOLAR (VOLUNTEER ARMY)

ACTIONS AT FORT BENNING ¹

T. O. Jacobs

Human Resources Research Organization

and

Daniel L. Rickett

United States Army Infantry Human Research Unit

An evaluation of the first year of experience at Fort Benning with actions designed to increase attractiveness of military service and thus decrease reliance on inductions (VOLAR). Through the use of a pre-VOLAR questionnaire, for baseline purposes, and periodic subsequent administrations, it was possible to assess VOLAR impact on career intentions and general attitudes toward the Army. VOLAR actions had greatest impact on soldiers' feelings about inequities, and less on needs for effective leadership, security, and pride in service. Soldiers' measured career intentions have increased systematically during the period of evaluation.

In October 1970, the U. S. Army initiated planning to achieve, eventually, zero reliance on the draft, selecting Fort Benning to be one of the first VOLAR posts. The objective was to conduct a quasi-experiment to formulate ways of achieving VOLAR objectives. The strategy was to identify effective changes at each installation, and then to export such changes throughout the Army. Clearly, evaluation of individual post actions had to be an essential element of this strategy, and the evaluation of Fort Benning VOLAR actions was given a correspondingly high priority by the Commanding General of the U. S. Army Infantry Center.

Method

The evaluation philosophy called for identifying the dependent variables significant to the VOLAR program, and then developing a means for assessing the degree of change in these variables that could be attributed to VOLAR actions. VOLAR objectives are to improve Army life and professionalism, in order to increase accessions and retentions, on the one hand, and to improve attitudes toward the Army even among separatees, on the other hand.

This second objective was judged significant because those soldiers who decide to separate after the first enlistment provide word-of-mouth "advertising" which is quite important to the accomplishment of long-range objectives of the VOLAR program.

The evaluation at Fort Benning therefore used an attitude questionnaire as a primary tool, together with a variety of supplemental techniques (interviews of re-enlistees and separatees, collection of objective data on re-enlistment rates, AWOL rates, etc.).

The questionnaire dealt with:

1. Awareness of and satisfaction with specific VOLAR and MVA actions at Fort Benning.
2. Attitudes toward the Army in general.
3. Evaluation of Fort Benning as a military post.
4. Expressed intentions to re-enlist in the Army, or to separate at the end of the current tour.

This questionnaire was administered initially in late November 1970, just prior to implementation of VOLAR actions. It now has been administered a total of three additional times, to a total of 6,559 enlisted and commissioned personnel at Fort Benning, and 2,584 enlisted and commissioned personnel from a control post during the last half of FY 71 (two administrations).

Analyses of the questionnaire data consisted of comparison of responses from subsequent administrations with data obtained in the original baseline administration. Some details:

1. Attitudes toward specific VOLAR actions. General linear solutions on ANOVAs were performed on each item representing a specific VOLAR action, e.g., the use of civilian hires to replace enlisted KPs. For the fourth administration, the primary focus of this paper, data from a control post were not available. The variables consequently were time of administration (November 1970 vs November 1971), grade (officer vs enlisted), and tour (first vs extended), with the time main effect as the primary criterion of significance.

2. General attitudes toward military service. Seventy items were included in a general pool of attitude statements. A principal components factor analysis with a VARIMAX rotation was conducted for officer and enlisted first-tour personnel separately on the data from the second administration to test generalizability of the factor

structures . Surprisingly, there was almost total correspondence in the factor structure obtained for each group. Factor composite scores consequently were obtained, using one standard set of items for each factor, weighting each equally. These were subjected to the general linear solution ANOVAs used for specific VOLAR action items. In addition, specific VOLAR actions were regressed against the attitude composites, and against expressed career intentions, as were the factor composite scores themselves.

3. Career intentions. ANOVAs were conducted on this single item at all administrations, in addition to the regression analyses just described.

Results

This paper will focus primarily on some of the results from the last data collection point (November 1971), which closes the first full year of evaluation.

General Attitudes Toward the Army

As was noted earlier, the survey instrument contained 70 items concerning general attitudes toward the Army. Factor analysis of these general items yielded either four or five factors, regardless of what was used as a diagonal entry, with *both* first tour groups (officers and enlisted), *over different administrations*. The only difference between the four and five factor solution was that, depending on the diagonal entry, one group of items either loaded negatively on an earlier factor, or emerged as a separate factor. Because they replicated so well, both seemed to be powerful solutions. However, the four factor solution was chosen, as somewhat more meaningful. Its factors could easily be interpreted as needs which must be satisfied if a military career is to be attractive:

1. Pride in military service.
2. Security--physical and psychological.
3. Equitable rewards in exchange for performance of duty.
4. Competent and understanding leadership.

Correlations among the factor composite scores² and between them and career intentions for both enlisted and commissioned first tour personnel are shown in Table 1, separately by category of respondents, and time of survey. The enlisted personnel who expressed more positive career intentions were those who felt more involvement (pride) in military service and a greater feeling of security. The same holds true for commissioned respondents, though

Table 1
Correlations Among Factor Composite Scores
and Career Intentions

	Enlisted				Commissioned			
	Involv	Sec	Rew	Lead	Involv	Sec	Rew	Lead
Career Intentions	.42 (.36)*	.74 (.49)	.17 (.21)	.23 (.24)	.58 (.38)	.70 (.53)	.48 (.40)	.31 (.32)
Involvement		.83 (.80)	.29 (.32)	.54 (.56)		.87 (.78)	.43 (.35)	.52 (.49)
Security			.21 (.35)	.58 (.56)			.49 (.44)	.47 (.51)
Reward				.18 (.31)				.36 (.25)

*Parenthetical entries obtained from June 1971 data

much more substantially for the security factor. Further, while the relationships between the factor scores and career intentions did not change over a period of six months for enlisted personnel, the relationship between both pride and security, and career intentions became much stronger for commissioned personnel. It might be *inferred* that these need areas may have become more salient for these first tour officer personnel during that period of time, while not changing for enlisted.

Change in average level of response on these factor composite scores was also examined for the November 1970-November 1971 time period. For first-tour personnel, the only area in which significant positive change was *not* found was pride in service. Without an appropriate control group, it is, of course, not possible to attribute the significant changes in the other three to VOLAR actions; however, the inference that VOLAR actions did produce these changes is certainly possible.

Individual VOLAR Actions

Attitudes of satisfaction toward 64 separate VOLAR actions have now been followed for the first year of VOLAR. Of these 64, satisfaction has increased significantly on 62. Perhaps more impressive, however, than the *number* of significant changes is the magnitude of observed changes. Using the time main effect as a criterion, the range of f-ratios was from .150 (NS) to 535.001. The average f-ratio was 131.3, and 31 were over 100. These significance levels are quite impressive, exceeding levels normally found in such studies.

The specific VOLAR actions showing the greatest mean change from November 1970 to November 1971 were, for enlisted personnel:

1. Compensatory time off during the week for...weekend details.
2. Frequency with which military personnel are required to perform kitchen police.
3. Policies on travel distance during off-duty time.
4. Opportunity to eat breakfast in the unit mess hall after sleeping late on weekends and holidays.
5. Policy concerning beer in the barracks.
6. Privacy and individuality in troop barracks.
7. Frequency with which military personnel are required to perform refuse and garbage pick-up details.
8. Frequency with which military personnel are required to cut grass and police the post.

For officer personnel, actions showing the greatest mean change were:

1. Privacy and individuality in troop barracks.
2. Frequency with which military personnel are required to perform kitchen police.
3. Policies on travel distance during off-duty time.
4. Policies and regulations affecting OBV-2 officers (regarding purchase of the Army blue uniform and its wear at social functions).
5. Frequency with which military personnel are required to cut grass and police the post.

The big impact items for enlisted personnel were those which replaced the soldier with a civilian hire for menial details not really a part of his MOS, and actions that gave him more individuality and freedom. For first tour officers, big impact items tended to be those which were big for their men.

Expressed Career Intentions

Respondents also were asked to choose between the four alternatives shown in Table 2, concerning their intentions to remain in the Army. As can be seen, there has been a steady increase in favorable responses to this item.

Table 2

Percentages of Enlisted and Commissioned First Tour Personnel
Responding to Career Intentions Alternatives

First Tour	Remain to Retirement	A While Longer	Undecided	Leave the Army
Enlisted				
Nov 70	1.4	4.9	12.0	81.6
June 71	2.1	7.5	13.3	77.1
Nov 71	3.6	10.7	17.7	68.0
Commissioned				
Nov 70	7.3	14.3	20.1	58.3
Jun 71	10.1	12.9	14.3	62.7
Nov 71	12.1	20.9	16.5	50.5

The change from November 1970 to November 1971 is highly significant for first tour soldiers ($p < .001$), and approaches significance for first tour officers ($.05 < p < .10$). Changes for extended tour groups did not approach significance (and are not shown); their average responses were quite high initially.

Actual Re-enlistment Experience

While more positive attitudes toward the Army were a major objective of VOLAR actions, the primary objective, of course, was actual re-enlistment. The evaluation therefore included analysis of actual re-enlistment experience at Fort Benning over the period of

the VOLAR experiment, in comparison with previous experience. For enlisted personnel, there was a positive change in rate of enlistments per thousand operating strength, from 9.70 to 10.27, a change which failed to reach statistical significance. However, it should be noted that *requirements* for re-enlistment qualification were made substantially more stringent during the last part of CY 1971; otherwise, the increase might well have been larger, and significant. That this might well be true is indicated by examination of "re-enlistment" experience for officers. While the expressed career intentions of first tour officers did not become significantly more positive, their actual behavior did. The number of OBV officers requesting extension or RA appointment was very significantly higher ($p < .001$) during CY 1971 than during CY 1970.

Interview Data with Re-enlistees and Separatees

One of the questions asked in interviews of a sample of both re-enlistees and separatees was the influence Project VOLAR had had on their decision either to remain in the Army or to separate. Table 3 shows the results for 68 re-enlistees and 114 separatees of all rank levels. The difference in distribution of responses was significant (Chi-square) at the .03 level. This Table also provides a means of inferring the influence of Project VOLAR at Fort Benning on actual re-enlistments.

Table 3
Influence of VOLAR on Decision

	Strong to Stay	Some to Stay	No Infl.	Some to Leave	Strong to Leave
Re-enlistees	11	13	40	1	3
Separatees	4	20	80	6	4
Chi-square = 10.87			$p < .03$		

Discussion

It would appear that certain conclusions can be drawn from the Fort Benning VOLAR evaluation, even though it could not be a rigorously controlled experiment. First, there can be little doubt that the surveyed populations at Benning were keenly aware of the VOLAR actions. Further, general attitudes toward the Army have improved during this

period, and expressed career intentions show significant change in the desired direction. While evidence that VOLAR actions have caused these changes is only circumstantial, it is inviting to infer that they have, though the effect may not have been as large as might have been desired.

Given this, other observations can also be made concerning MVA. in perspective. Most of the high impact items at Fort Benning were directed at reducing the inequities of military life, especially for the enlisted man. While these actions were apparently urgently needed, they fall into the Herzberg category of hygiene factors. Little has been done with motivator variables, though theoretically these are the variables that *should* enhance the attractiveness of a service career. One possible reason for this lack of attention to motivator variables is that such changes are generally quite difficult to implement within formal organizations especially at the lower operating levels. However, at the same time, this may very well be the *required* next direction for VOLAR actions to take.

Footnotes

¹Research reported in this paper was performed at HumRRO Division No. 4, Fort Benning, Georgia, under Department of the Army contract; the contents of this paper do not necessarily reflect official opinions or policies of the Department of the Army. Reproduction in whole or in part is permitted for any purpose of the United States Government.

²These composite scores were calculated by assigning unity weights to items loading higher than .50 on a factor. One result was that the factors were no longer orthogonal.

Reference

Herzberg, F., Mausner, B., & Snyderman, B. A. *The Motivation to Work*. New York: John Wiley, 1959.